

Singularity structures and impacts on parameter estimation in finite mixtures of distributions

Nhat Ho and XuanLong Nguyen

Department of Statistics
University of Michigan

Abstract

Singularities of a statistical model are the elements of the model's parameter space which make the corresponding Fisher information matrix degenerate. These are the points for which estimation techniques such as the maximum likelihood estimator and standard Bayesian procedures do not admit the root- n parametric rate of convergence. We propose a general framework for the identification of singularity structures of the parameter space of finite mixtures, and study the impacts of the singularity levels on minimax lower bounds and rates of convergence for the maximum likelihood estimator over a compact parameter space. Our study makes explicit the deep links between model singularities, parameter estimation convergence rates and minimax lower bounds, and the algebraic geometry of the parameter space for mixtures of continuous distributions. The theory is applied to establish concrete convergence rates of parameter estimation for finite mixture of skewnormal distributions. This rich and increasingly popular mixture model is shown to exhibit a remarkably complex range of asymptotic behaviors which have not been hitherto reported in the literature. ¹

1 Introduction

In the standard asymptotic theory of parametric estimation, a customary regularity assumption is the non-singularity of the Fisher information matrix defined by the statistical model (see, for example, Lehmann and Casella [1998] (pg. 124); or van der Vaart [1998], Sec. 5.5). This condition leads to the cherished root- n consistency, and in many cases the asymptotic normality of parameter estimates. When the non-singularity condition fails to hold, that is, when the true parameters represent a singular point in the statistical model, very little is known about the asymptotic behavior of their estimates.

The singularity situation might have been brushed aside as idiosyncratic by some parametric statistical modelers in the past. As complex and high-dimensional models are increasingly embraced by statisticians and practitioners alike, singularities are no longer a rarity — they start to take a highly visible place in modern statistics. For example, the many zeros present in a high-dimensional linear regression problem represent a type of singularities of the underlying model, points corresponding to rank-deficient Fisher information matrices [Hastie et al., 2015]. In another example, the zero skewness in the family of skewed distributions represents a singular point [Chiogna, 2005]. In both examples, singularity points are quite easy to spot out — it is the impacts of their presence on improved parameter estimation procedures and the asymptotic properties such procedures entail that are nontrivial matters occupying the best efforts of many researchers in the past decade. The textbooks by Bühlmann and van de Geer [2011], Hastie et al. [2015], for instance, address such issues for high-dimensional regression problems, while the recent papers by Ley and Paindaveine [2010],

¹This research is supported in part by grants NSF CAREER DMS-1351362 and NSF CNS-1409303 to XN. The authors would like to acknowledge Larry Wasserman and Ya'acov Ritov for valuable discussions related to this work.

AMS 2000 subject classification: Primary 62F15, 62G05; secondary 62G20.

Keywords and phrases: Fisher singularities, system of polynomial equations, semialgebraic set, mixture model, non-linear partial differential equation, minimax lower bound, maximum likelihood estimation, convergence rates, Wasserstein distances.

Hallin and Ley [2012, 2014] investigate statistical inference in the skewed families for distribution. By contrast, with finite mixture models — a popular and rich class of modeling tools for density estimation and heterogeneity inference [Lindsay, 1995] and a subject of this paper, the singularity phenomenon is not quite well understood, to the best of our knowledge, except for specific instances.

One of the simplest instances is the singularity of Fisher information matrix in an (overfitted) finite mixture that includes a homogeneous distribution. Lee and Chesher [1986] analyzed a test of heterogeneity based on finite mixtures, addressing the challenge arising from the aforementioned singularity. Recent works on the related topic include Chen and Chen [2003], Kasahara and Shimotsu [2014]. Rotnitzky et al. [2000] investigated likelihood-based parameter estimation in a somewhat general parametric modeling framework, subject to the constraint that the Fisher information matrix is one rank deficient. For overfitted finite mixtures, Chen [1995] showed that under a condition of strong identifiability, there are estimators which achieve the generic convergence rate $n^{-1/4}$ for parameter estimation. Recent works also established generic behaviors of estimation under somewhat broader settings of overfitted finite mixture models with both maximum likelihood estimation and Bayesian estimation [Rousseau and Mengersen, 2011, Nguyen, 2013, Ho and Nguyen, 2016b].

The family of mixture models is far too rich to submit to a uniform kind of behavior of parameter estimation. In fact, it was shown only recently that even classical models such as the location-scale Gaussian mixtures, and the shape-rate Gamma mixtures, do not admit such a generic rate of convergence for an estimation method such as MLE [Ho and Nguyen, 2016a]. For instance, singularities arise in the finite mixtures of Gamma distributions, even when the number of mixing components is known — this phenomenon results in an extremely slow convergence behavior for the model parameters lying in the vicinity of singular points, even though such parameters are (perfectly) identifiable. Finite mixtures of Gaussian distributions, though identifiable, exhibit both minimax lower bounds and maximum likelihood estimation rates that are directly linked to the solvability of a system of real polynomial equations, rates which deteriorate quickly with the increasing number of extra mixing components. The results obtained for such specific instances contain considerable insights about parameter estimation in finite mixture models, but they only touch upon the surface of a more general phenomenon. Indeed, as we shall see there is a much richer spectrum of asymptotic behavior in which regular (non-singular) mixtures, strongly identifiable mixtures, and weakly identifiable mixture models (such as the one studied by Ho and Nguyen [2016a]) occupy but a small spot.

Objectives and main results In this paper we propose a theoretical framework for analyzing parameter estimation behavior in finite mixture models, addressing directly the situations where the non-singularity condition of the Fisher information matrix may not hold. Our approach is to take on a systematic investigation of the singularity structure of a compact and multi-dimensional parameter space of mixture models, and then study the impacts of the presence of singularities on parameter estimation. It is no longer sufficient to speak of the standard notion of Fisher information singularities. A more fundamental notion that we introduce is called *singularity level*, a natural or infinite value given to every element in the parameter space. Fisher information singularities simply correspond to points in the parameter space whose singularity level is non-zero. Within the set of Fisher information singularities the parameter space can be partitioned into disjoint subsets determined by different singularity levels. The singularity level describes in a precise manner the variation of the mixture likelihood with respect to changes in model parameters. This concept enables us to quantify the varying degrees of identifiability and singularity, some of which were implicitly exploited in previous works mentioned above.

The statistical implication of the singularity level is easy to describe: given an i.i.d. n -sample from a (true) mixture model, a parameter value of singularity level r admits $n^{-1/2(r+1)}$ minimax lower

bound for any estimator tending to the true parameter(s), as well as the same maximum likelihood estimator's convergence rate (up to a logarithmic factor and under some conditions). Thus, singularity level 0 results in root- n convergence rate for parameter estimation. Fisher singularity corresponds to singularity level 1 or greater than 1, resulting in convergence rates $n^{-1/4}, n^{-1/6}, n^{-1/8}$ or so on. The detailed picture of the distribution of singularity levels, however, can be extremely complex to capture. Remarkably, there are examples of finite mixtures for which the compact parameter space can be partitioned into disjoint subsets whose singularity level ranges from 0 to 1 to 2, ..., up to infinity. As a result, if we were to vary the true parameter values, we would encounter a phenomenon akin to that of "phase transition" on the statistical efficiency of parameter estimation occurring within the same model class.

Techniques A major component of our general framework is a procedure for characterizing subsets of points carrying the same singularity level. It will be shown that these points are in fact a subset of a real affine variety. A real affine variety is a set of solutions to a system of real polynomial equations. The polynomial equations can be derived explicitly by the kernel density functions that define a given mixture distribution. The study of the solutions of polynomial equations is a central subject of algebraic geometry [Stumfel, 2002, Cox et al., 2007]. The connections between statistical models and algebraic geometry have been studied for discrete Markov random fields [Drton et al., 2009], as well as finite mixtures of categorical data [Allman et al., 2009]. For finite mixtures of continuous distributions, the link to algebraic geometry is distilled from a new source of algebraic structure, in addition to the presence of mixing measures: it is traced to the partial differential equations satisfied by the mixture model's kernel density function. For Gaussian mixtures, it is the relation captured by Eq. (3) for the Gaussian kernel. The partial differential equations can be nonlinear, with coefficients given by rational functions defined in terms of model parameters. It is this relation that is primarily responsible for the complexity of the singularity structure. A quintessential example of such a relation is given by Eq. (2) for the skewnormal kernel densities.

Although our method for the analysis of singularity structure and the asymptotic theory for parameter estimation can be used to re-derive old and existing results such as those of Chen [1995], Ho and Nguyen [2016a], a substantial outcome is to establish new results on mixture models for which no asymptotic theory have hitherto been achieved. This leads us to a story of finite mixtures of skewnormal distributions. The skewnormal distribution was originally proposed in Azzalini [1986], Azzalini and Valle [1996], Azzalini and Capitanio [1999]. The skewnormal generalizes normal (Gaussian) distribution, which is enhanced by the capability of handling asymmetric (skewed) data distributions. Due to its more realistic modeling capability for multi-modality and asymmetric components, skewnormal mixtures are increasingly adopted in recent years for model based inference of heterogeneity by many researchers [Lin et al., 2007, Arellano-Valle et al., 2008, 2009, Lin, 2009, Schnatter and Pyne, 2009, Ghosal and Roy, 2011, Lee and McLachlan, 2013, Prates et al., 2013, Canale and Scarpa, 2015, Zeller et al., 2015]. Due to its usefulness, a thorough understanding of the asymptotic behavior of parameter estimation for skewnormal mixtures is also of interest in its own right.

The singularity structure of the skewnormal mixtures is perhaps one of the more complex among the parametric mixture models that we have typically encountered in the literature. By comparison, strongly identifiable models admit the same singularity level (1, to be precise) for all parameter values residing in a compact space, resulting in $n^{-1/4}$ convergence rate for the MLE. Most mixture models whose kernel density function has only one type of parameter, such as location mixtures or scale mixtures, are in this category. Location-scale Gaussian mixtures are a step up in the complexity, in that all their parameter values carry the same singularity level, which depends only on the number of extra

mixing components. Yet this is not the picture of skewnormal mixtures. We will be able to identify subsets with singularity level 0, 1, 2, ... all the way up to infinity. Even in the setting of mixtures with known number of mixing components, the singularity structure is remarkably complex. Thus, the results for skewnormal mixtures present an useful illustration for the full power of the general theory for finite mixtures of continuous distributions.

The source of complexity of skewnormal mixtures is the structure of the skewnormal kernel density. The evidence for the latter was already made clear by Chiogna [2005], Ley and Paindaveine [2010], Hallin and Ley [2012, 2014], whose works provided a thorough picture of the singularities for the class of skewnormal densities, and their impacts on the non-standard rates of convergence of MLE. Not only can we recover the results of Hallin and Ley [2012, 2014] in terms of rates of convergence, which correspond to a trivial “mixture” that has exactly one skewnormal component, an entirely new set of results are established for mixtures of two or more components. It is in this setting that new types of singularities arise out of the interactions between distinct skewnormal components. These interactions define the subset of singular points of a given level that can be characterized by a system of real polynomial equations. This algebraic geometric characterisation allows us to establish either the precise singularity level or an upper bound for the mixture model’s entire parameter space.

The plan for the remainder of our paper is as follows. Section 2 lays out the notations and relevant concepts such as parameter spaces and the underlying geometries. Section 3 presents the general framework of analysis of singularity structure, and the impact on convergence rates of parameter estimation for singular points of a given singularity level. Section 4 and Section 5 illustrate the theory on the finite mixture of skewnormal distributions, by giving concrete minimax bounds and MLE convergence rates for this class of models for the first time. We conclude with a discussion in Section 6. Further details of the proofs and some additional results are given in the Appendices.

2 Background

A finite mixture of continuous distributions admits density of the form $p_G(x) = \int f(x|\eta)dG(\eta)$ with respect to Lebesgue measure on an Euclidean space for x , where $f(x|\eta)$ denotes a probability density kernel, η is a multi-dimensional parameter taking values in a subset of an Euclidean space Θ , G denotes a discrete mixing distribution on Θ . The number of support points of G represents the number of mixing components in mixture model. Suppose that $G = \sum_{i=1}^k p_i \delta_{\eta_i}$, then $p_G(x) = \sum_{i=1}^k p_i f(x|\eta_i)$.

2.1 Parameter spaces and geometries

There are different kinds of parameter space and geometries that they carry which are relevant to our work. We proceed to describe them in the following.

Natural parameter space The customarily defined parameter space of the k -mixture of distributions is that of the mixing component parameters η_i , and mixing probabilities p_i . Throughout this paper, it is assumed that $\eta_i \in \Theta$, which is a compact subset of \mathbb{R}^d for some $d \geq 1$, for $i = 1, \dots, k$. The mixing probability vector $\mathbf{p} = (p_1, \dots, p_k) \in \Delta^{k-1}$, the $(k-1)$ -probability simplex. To simplify the theory we will further assume (in Section 4) that all $p_i \geq c_0$ for some small positive constant c_0 . For the remainder of the paper, we also use Ω to denote the compact subset of the Euclidean space for parameters $(\mathbf{p}, \boldsymbol{\eta})$.

Example 2.1. The skewnormal density kernel on the real line has three parameters $\eta = (\theta, \sigma, m) \in$

$\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$, namely, the location, scale and skewness (shape) parameters. It is given by, for $x \in \mathbb{R}$,

$$f(x|\theta, \sigma, m) := \frac{2}{\sigma} f\left(\frac{x - \theta}{\sigma}\right) \Phi(m(x - \theta)/\sigma),$$

where $f(x)$ is the standard normal density and $\Phi(x) = \int f(t)1(t \leq x) dt$. This generalizes the Gaussian density kernel, which corresponds to fixing $m = 0$. The parameter space for the k -mixture of skewnormals is therefore a subset of an Euclidean space for the mixing probabilities p_i and mixing component parameters $\eta_i = (\theta_i, v_i = \sigma_i^2, m_i) \in \mathbb{R}^3$. For each $i = 1, \dots, k$, θ_i, σ_i, m_i are restricted to reside in compact subsets $\Theta_1 \subset \mathbb{R}, \Theta_2 \subset \mathbb{R}_+, \Theta_3 \subset \mathbb{R}$ respectively, i.e., $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$.

Semialgebraic sets The singularity structure of the parameter space carries a different geometry. It will be described in terms of the zero sets (sets of solutions) of systems of real polynomial equations. The zero set of a system of real polynomial equations is called a (real) affine variety [Cox et al., 2007]. In fact, the sets which describe the singularity structure of finite mixtures are not affine varieties per se. We will see that they are the intersection between real affine varieties – the real-valued solutions of a finite collection of equations of the form $P(\mathbf{p}, \boldsymbol{\eta}) = 0$, and the set of parameter values satisfying $Q(\mathbf{p}, \boldsymbol{\eta}) > 0$, for some real polynomials P and Q . The intersection of these sets is also referred to as semialgebraic sets.

Example 2.2. Continuing on the example of skewnormal mixtures, we will see that first two types of singularities that arise from the mixture of skewnormals are solutions of the following polynomial equations

(i) Type A: $P_1(\boldsymbol{\eta}) = \prod_{j=1}^k m_j$.

(ii) Type B: $P_2(\boldsymbol{\eta}) = \prod_{1 \leq i \neq j \leq k} \left\{ (\theta_i - \theta_j)^2 + \left[\sigma_i^2(1 + m_j^2) - \sigma_j^2(1 + m_i^2) \right]^2 \right\}$.

These are just two among many more polynomials and types of singularities that we will be able to enumerate in the sequel. We quickly note that Type A refers to the one (and only) kind of singularity intrinsic to the skewnormal kernel: $P_1 = 0$ if either one of the $m_j = 0$ — one of the skewnormal components is actually normal (symmetric). This type of singularity has received in-depth treatments by a number of authors [Chiogna, 2005, Ley and Paindaveine, 2010, Hallin and Ley, 2012, 2014]. On the other hand, Type B refers to something intrinsic to a mixture model, as it describes the relation of parameters of distinct mixing components i and j .

Space of mixing measures and transportation distance As described in the Introduction, a goal of this work is to turn the knowledge about the nature of singularities of parameter space Ω into the statistical efficiency of parameter estimation procedures. For this purpose, the convergence of parameters in a mixture model is most naturally analyzed in terms of the convergence in the space of mixing measures endowed by transportation distance (Wasserstein distance) metrics [Nguyen, 2013]. This is because the role played by parameters $\mathbf{p}, \boldsymbol{\eta}$ in the mixture model is via mixing measure G . It is mixing measure G that determines the mixture density p_G according to which the data are drawn from. Since the map $(\mathbf{p}, \boldsymbol{\eta}) \mapsto G(\mathbf{p}, \boldsymbol{\eta}) = G = \sum p_i \delta_{\eta_i}$ is many-to-one, we shall treat a pair of parameter vectors $(\mathbf{p}, \boldsymbol{\eta}) = (p_1, \dots, p_k; \eta_1, \dots, \eta_k)$ and $(\mathbf{p}', \boldsymbol{\eta}') = (p'_1, \dots, p'_{k'}; \eta'_1, \dots, \eta'_{k'})$ to be equivalent if the corresponding mixing measures are equal, $G(\mathbf{p}, \boldsymbol{\eta}) = G(\mathbf{p}', \boldsymbol{\eta}')$.

For $r \geq 1$, the Wasserstein distance of order r between $G(\mathbf{p}, \boldsymbol{\eta})$ and $G(\mathbf{p}', \boldsymbol{\eta}')$ takes the form (cf. Villani [2003]),

$$W_r(G(\mathbf{p}, \boldsymbol{\eta}), G(\mathbf{p}', \boldsymbol{\eta}')) = \left(\inf \sum_{i,j} q_{ij} \|\eta_i - \eta'_j\|_r^r \right)^{1/r},$$

where $\|\cdot\|_r$ is the ℓ_r norm endowed by the natural parameter space, the infimum is taken over all couplings \mathbf{q} between \mathbf{p} and \mathbf{p}' , i.e., $\mathbf{q} = (q_{ij})_{ij} \in [0, 1]^{k \times k'}$ such that $\sum_{i=1}^{k'} q_{ij} = p_j$ and $\sum_{j=1}^k q_{ij} = p'_i$ for any $i = 1, \dots, k$ and $j = 1, \dots, k'$. (For the example of skewnormal mixtures, if $\boldsymbol{\eta} = (\theta, v, m)$ and $\boldsymbol{\eta}' = (\theta', v', m')$, then $\|\boldsymbol{\eta} - \boldsymbol{\eta}'\|_r^r := |\theta - \theta'|^r + |v - v'|^r + |m - m'|^r$).

Suppose that a sequence of probability measures $G_n = \sum_i p_i^n \delta_{\boldsymbol{\eta}_i^n}$ tending to G_0 under W_r metric at a rate $\omega_n = o(1)$. If all G_n have the same number of atoms $k_n = k_0$ as that of G_0 , then the set of atoms of G_n converge to the k_0 atoms of G_0 , up to a permutation of the atoms, at the same rate ω_n under $\|\cdot\|$. If G_n have the varying $k_n \in [k_0, k]$ number of atoms, where k is a fixed upper bound, then a subsequence of G_n can be constructed so that each atom of G_0 is a limit point of a certain subset of atoms of G_n — the convergence to each such limit also happens at rate ω_n . Some atoms of G_n may have limit points that are not among G_0 's atoms — the total mass associated with those “redundant” atoms of G_n must vanish at the generally faster rate ω_n^r .

2.2 Estimation settings

The impact of singularities on parameter estimation behavior is dependent on whether the mixture model is fitted with a known number of mixing components, or if only an upper bound on the number of mixing components is known. The former model fitting setting is called “e-mixtures” for short, while the latter “o-mixtures” (“e” for exact-fitted and “o” for over-fitted).

Specifically, given an i.i.d. n -sample X_1, X_2, \dots, X_n according to the mixture density $p_{G_0}(x) = \int f(x|\boldsymbol{\eta}) G_0(d\boldsymbol{\eta})$, where $G_0 = G(\mathbf{p}^0, \boldsymbol{\eta}^0) = \sum_{i=1}^{k_0} p_i^0 \delta_{\boldsymbol{\eta}_i^0}$ is unknown mixing measure with exactly k_0 distinct support points. We are interested in fitting a mixture of k mixing components, where $k \geq k_0$, using the n -sample X_1, \dots, X_n . In the e-mixture setting, $k = k_0$ is known, so an estimate G_n (such as the maximum likelihood estimate) is drawn from ambient space \mathcal{E}_{k_0} , the set of probability measures $G = G(\mathbf{p}, \boldsymbol{\eta})$ with exactly k_0 support points, where $(\mathbf{p}, \boldsymbol{\eta}) \in \Omega$. In the o-mixture setting, \hat{G}_n is drawn from ambient space \mathcal{O}_k , the set of probability measures $G = G(\mathbf{p}, \boldsymbol{\eta})$ with at most k support points, where $(\mathbf{p}, \boldsymbol{\eta}) \in \Omega$.

Assumption Throughout this paper, several conditions on the kernel density $f(x|\boldsymbol{\eta})$ are assumed to hold. Firstly, the collection of kernel densities f as $\boldsymbol{\eta}$ varies is linearly independent. It follows that the mixture model is identifiable, i.e., $p_G(x) = p_{G_0}(x)$ for almost all x entails $G = G_0$. Secondly, we say $f(x|\boldsymbol{\eta})$ satisfies a uniform Lipschitz condition of order r , for some $r \geq 1$, if f as a function of $\boldsymbol{\eta}$ is differentiable up to order r , and that the partial derivatives with respect to $\boldsymbol{\eta}$, namely $\partial^{|\boldsymbol{\kappa}|} f / \partial \boldsymbol{\eta}^{\boldsymbol{\kappa}}$, for any $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_d) \in \mathbb{N}^d$ such that $|\boldsymbol{\kappa}| := \kappa_1 + \dots + \kappa_d = r$ satisfy the following: for any $\boldsymbol{\gamma} \in \mathbb{R}^d$,

$$\sum_{|\boldsymbol{\kappa}|=r} \left| \left(\frac{\partial^{|\boldsymbol{\kappa}|} f}{\partial \boldsymbol{\eta}^{\boldsymbol{\kappa}}} (x|\boldsymbol{\eta}_1) - \frac{\partial^{|\boldsymbol{\kappa}|} f}{\partial \boldsymbol{\eta}^{\boldsymbol{\kappa}}} (x|\boldsymbol{\eta}_2) \right) \right| \boldsymbol{\gamma}^{\boldsymbol{\kappa}} \leq C \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|_r^\delta \|\boldsymbol{\gamma}\|_r^r$$

for some positive constants δ and C independent of x and $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \Theta$. It is simple to verify that most kernel densities used in mixture modeling, including the skewnormal kernel, satisfy the uniform Lipschitz property over compact domain Θ , for any finite $r \geq 1$.

Notation We utilize several familiar notions of distance for mixture densities, with respect to Lebesgue measure μ . They are total variation distance $V(p_G, p_{G_0}) = \frac{1}{2} \int |p_G(x) - p_{G_0}(x)| d\mu(x)$ and Hellinger distance $h^2(p_G, p_{G_0}) = \frac{1}{2} \int \left(\sqrt{p_G(x)} - \sqrt{p_{G_0}(x)} \right)^2 d\mu(x)$.

3 Singularity structure in finite mixture models

3.1 Beyond Fisher information

Given a mixture model

$$\left\{ p_G(x) \middle| G = G(\mathbf{p}, \boldsymbol{\eta}) = \sum_{i=1}^k p_i \delta_{\eta_i}, (\mathbf{p}, \boldsymbol{\eta}) \in \Omega \right\}$$

from some given finite k and f a given kernel density (e.g., skewnormal), let l_G denote the score vector, that is, the column vector made of the partial derivatives of the log-likelihood function $\log p_G(x)$ with respect to each of the model parameters represented by $(\mathbf{p}, \boldsymbol{\eta})$. The Fisher information matrix is then given by $I(G) = \mathbb{E}(l_G l_G^\top)$, where the expectation is taken with respect to p_G . We say that the parameter vector $(\mathbf{p}, \boldsymbol{\eta})$ (and the corresponding mixing measure $G = G(\mathbf{p}, \boldsymbol{\eta})$) is a singular point in the parameter space (resp., ambient space of mixing measures), if $I(G)$ is degenerate. Otherwise, $(\mathbf{p}, \boldsymbol{\eta})$ (resp., G) is a non-singular point.

According to the standard asymptotic theory, if the true mixing measure G_0 is non-singular, *and* the number of mixing components $k_0 = k$ (that is, we are in the e-mixture setting), then basic estimators such as the MLE or Bayesian estimator yield the optimal root- n rate of convergence. Although the standard theory remains silent when $I(G_0)$ is degenerate, it is clear that the root- n rate may no longer hold. Moreover, there may be a richer range of behaviors for parameter estimation, requiring us to look into the deep structure of the zeros of $I(G_0)$. This will be our story for both settings of e-mixtures and o-mixtures. In fact, the Fisher information matrix $I(G_0)$ is no longer sufficient in assessing parameter estimation behaviors.

Example 3.1. To illustrate in the context of skewnormal mixtures, where parameter $\boldsymbol{\eta} = (\theta, v, m)$, observe that the mixture density structure allows the following characterization: $I(G)$ is degenerate if and only if the collection of partial derivatives

$$\left\{ \frac{\partial p_G(x)}{\partial p_j}, \frac{\partial p_G(x)}{\partial \eta_j} \right\} := \left\{ \frac{\partial p_G(x)}{\partial p_j}, \frac{\partial p_G(x)}{\partial \theta_j}, \frac{\partial p_G(x)}{\partial v_j}, \frac{\partial p_G(x)}{\partial m_j} \middle| j = 1, \dots, k \right\}$$

as functions of x are not linearly independent. This is equivalent to having that for some coefficients (α_{ij}) , $i = 1, \dots, 4$ and $j = 1, \dots, k$, not all of which are zeros, there holds

$$\sum_{j=1}^k \alpha_{1j} f(x|\eta_j) + \alpha_{2j} \frac{\partial f}{\partial \theta}(x|\eta_j) + \alpha_{3j} \frac{\partial f}{\partial v}(x|\eta_j) + \alpha_{4j} \frac{\partial f}{\partial m}(x|\eta_j) = 0, \quad (1)$$

for almost all $x \in \mathbb{R}$. Lemma 4.1 later shows that the (Fisher information matrix's) singular points are the zeros of some polynomial equations.

We have seen that for the e-mixtures G is non-singular if the collection of density kernel functions $f(x|\eta)$ and their first partial derivatives with respect to each model parameter are linearly independent. This condition is also known as the first-order identifiability. For o-mixtures, the relevant notion is the

second-order identifiability [Chen, 1995, Nguyen, 2013, Ho and Nguyen, 2016b]: the condition that the collection of kernel density functions $f(x|\eta)$, their first and second partial derivatives, are linearly independent. This condition fails to hold for skewnormal kernel densities, whose first and second partial derivatives are linked by the following nonlinear partial differential equations:

$$\begin{cases} \frac{\partial^2 f}{\partial \theta^2} - 2 \frac{\partial f}{\partial v} + \frac{m^3 + m}{v} \frac{\partial f}{\partial m} = 0. \\ 2m \frac{\partial f}{\partial m} + (m^2 + 1) \frac{\partial^2 f}{\partial m^2} + 2vm \frac{\partial^2 f}{\partial v \partial m} = 0. \end{cases} \quad (2)$$

The proof of these identities can be found in Lemma 8.1 in Appendix B. Note that if $m = 0$, the skewnormal kernel becomes normal kernel, which admits a (simpler) linear relationship:

$$\frac{\partial^2 f}{\partial \theta^2} = 2 \frac{\partial f}{\partial v}. \quad (3)$$

This relation plays a fundamental role in the analysis of finite mixtures of location-scale normal distributions [Ho and Nguyen, 2016a]. Compared to Gaussian density kernel, the nonlinear relationship exhibited by skewnormal density kernel results in a much richer behavior. Analyzing this requires the development of a more general theory that we now embark on.

3.2 Behavior of likelihood in a Wasserstein neighborhood

Instead of dwelling on the Fisher information matrix, we shall employ a direct approach which studies the behavior of the likelihood function $p_G(x)$ as G varies in a Wasserstein neighborhood of a mixing measure $G_0 = \sum_{i=1}^{k_0} p_i^0 \delta_{\eta_i^0}$.

Fix $r \geq 1$, and consider a sequence of $G_n \in \mathcal{O}_k$, such that $W_r(G_0, G_n) \rightarrow 0$. Let $k_n \leq k$ be the number of distinct support points of G_n . Then each supporting atom η_i^0 as $i \in \{1, \dots, k_0\}$ of G_0 will have at least one atom of G_n that converges to. By relabelling the support points of G_n , we can express it as

$$G_n = \sum_{i=1}^{k_0} \sum_{j=1}^{s_i^n} p_{ij}^n \delta_{\eta_{ij}^n}, \quad (4)$$

where $\eta_{ij}^n \rightarrow \eta_i^0$ for all $i = 1, \dots, k_0$, $j = 1, \dots, s_i^n$. Additionally, $\sum_{i=1}^{k_0} s_i^n = k_n$. There exists a subsequence of G_n according to which k_n and all s_i^n are constant in n . (Note that for the setting of e-mixtures, the sequence of elements G_n is restricted to \mathcal{E}_{k_0} , so $k_n = k_0$ for all n . It follows that $s_i^n = 1$ for all $i = 1, \dots, k_0$. For o-mixtures, to simplify the presentation, we have omitted the cases where some G_n may have atoms that do not converge to the atoms of G_0). Thus, from here on we replace the sequence of G_n by this subsequence. To simplify the notation, n will be dropped from the superscript when the context is clear. In addition, we use the notation $\Delta \eta_{ij} := \eta_{ij} - \eta_i^0$ for $i = 1, \dots, k_0$, $j = 1, \dots, s_i$. Also, $p_{i.} := \sum_{j=1}^{s_i} p_{ij}$, and $\Delta p_i := p_{i.} - p_i^0$, for $i = 1, \dots, k_0$. (For e-mixtures, since $s_i = 1$ for all i , the notation is simplified further: let $\Delta \eta_i := \Delta \eta_{i1} = \eta_i - \eta_i^0$, $\Delta p_i = \Delta p_{i.} = p_i - p_i^0$ for all $i = 1, \dots, k_0$.) The following lemma relates Wasserstein distance metric to a semipolynomial of degree r (a semipolynomial is a polynomial of the absolute value of some variables).

Lemma 3.1. *Fix $r \geq 1$. For any element G represented by Eq. (4), define*

$$D_r(G_0, G) := \sum_{i=1}^{k_0} \sum_{j=1}^{s_i} p_{ij} \|\Delta \eta_{ij}\|_r^r + \sum_{i=1}^{k_0} |\Delta p_i|.$$

We have that $W_r^r(G, G_0) \asymp D_r(G_0, G)$, as $W_r(G_0, G) \downarrow 0$.

To investigate the behavior of likelihood function $p_G(x)$ as G tends to G_0 in Wasserstein distance W_r , by representation (4), we can express

$$p_G(x) - p_{G_0}(x) = \sum_{i=1}^{k_0} \sum_{j=1}^{s_i} p_{ij} (f(x|\eta_{ij}) - f(x|\eta_i^0)) + \sum_{i=1}^{k_0} \Delta p_i f(x|\eta_i^0). \quad (5)$$

By Taylor expansion up to order r , we obtain

$$p_G(x) - p_{G_0}(x) = \sum_{i=1}^{k_0} \sum_{j=1}^{s_i} p_{ij} \sum_{|\kappa|=1}^r \frac{(\Delta\eta_{ij})^\kappa}{\kappa!} \frac{\partial^{|\kappa|} f}{\partial \eta^\kappa}(x|\eta_i^0) + \sum_{i=1}^{k_0} \Delta p_i f(x|\eta_i^0) + R_r(x), \quad (6)$$

where $R_r(x)$ is the Taylor remainder. Moreover, it can be verified that $\sup_x |R_r(x)/W_r^r(G, G_0)| \rightarrow 0$ since f is uniform Lipschitz up to order r . We arrive at the following formulae, which measures the speed of change of the likelihood function as G varies in the Wasserstein neighborhood of G_0 :

$$\frac{p_G(x) - p_{G_0}(x)}{W_r^r(G, G_0)} = \sum_{|\kappa|=1}^r \sum_{i=1}^{k_0} \sum_{j=1}^{s_i} \left(\frac{p_{ij}(\Delta\eta_{ij})^\kappa / \kappa!}{W_r^r(G_0, G)} \right) \frac{\partial^{|\kappa|} f}{\partial \eta^\kappa}(x|\eta_i^0) + \sum_{i=1}^{k_0} \frac{\Delta p_i}{W_r^r(G_0, G)} f(x|\eta_i^0) + o(1). \quad (7)$$

The right hand side of Eq. (7) is a linear combination of the partial derivatives of f evaluated at G_0 . In addition, by Lemma 3.1, the coefficients of this linear representation is asymptotically equivalent to the ratio of two semipolynomials.

Equation (7) highlights the distinct roles of model parameters and the kernel density function in the formation of a mixture model's likelihood. The former appear only in the coefficients, while the latter provides the partial derivatives which appear as if basis functions for the linear combination. We wrote “as if”, because the partial derivatives of kernel f may not be linearly independent functions (recall the examples in Section 3.1). When a partial derivative of f can be represented as a linear combination of other partial derivatives, it can be eliminated from the expression in the right hand side. This reduction process may be repeatedly applied until all partial derivatives that remain are linearly independent functions. This motivates the following.

Definition 3.1. *The following representation is called r -minimal form of the mixture likelihood for a sequence of mixing measures G tending to G_0 in W_r metric:*

$$\frac{p_G(x) - p_{G_0}(x)}{W_r^r(G, G_0)} = \sum_{l=1}^{T_r} \left(\frac{\xi_l^{(r)}(G)}{W_r^r(G_0, G)} \right) H_l^{(r)}(x) + o(1), \quad (8)$$

which holds for all x , with the index l ranging from 1 to a finite T_r , if

- (1) $H_l^{(r)}(x)$ for all l are linearly independent functions of x , and
- (2) coefficients $\xi_l^{(r)}(G)$ are polynomials of the components of $\Delta\eta_{ij}$, and $\Delta p_i, p_{ij}$.

It is sufficient, but not necessary, to select functions $H_l^{(r)}$ from the collection of partial derivatives $\partial^{|\kappa|} f / \partial \eta^\kappa$ evaluated at particular atoms η_i^0 of G_0 , where $|\kappa| \leq r$, by adopting the elimination technique. The precise formulation of $\xi_l^{(r)}(G)$ and $H_l^{(r)}(x)$ will be determined explicitly by the specific G_0 . The r -minimal form for each G_0 is not unique, but they play a fundamental role in our notion of the singularity level of G_0 relative to a class of mixing distributions \mathcal{G} .

Definition 3.2. Fix $r \geq 1$ and let \mathcal{G} be a class of discrete probability measures which has a bounded number of support points in Θ . We say G_0 is r -singular relative to \mathcal{G} , if G_0 admits a r -minimal form given by Eq. (8), according to which there exists a sequence of $G \in \mathcal{G}$ tending to G_0 under W_r such that

$$\xi_l^{(r)}(G)/W_r^r(G, G_0) \rightarrow 0 \text{ for all } l = 1, \dots, T_r.$$

We now verify that the r -singularity notion is well-defined, in that it does not depend on any specific choice of the r -minimal form. This invariant property is confirmed by part (a) of the following lemma. Part (b) establishes a crucial monotonic property.

Lemma 3.2. (a) (Invariance) The existence of the sequence of G in the statement of Definition 3.2 holds for all r -minimal forms once it holds for at least one r -minimal form.

(b) (Monotonicity) If G_0 is r -singular for some $r > 1$, then G_0 is $(r - 1)$ -singular.

Proof. (a) The existence of the sequence of G described in the definition of a r -minimal form implies for that sequence, $(p_G(x) - p_{G_0}(x))/W_r^r(G, G_0) \rightarrow 0$ holds for any x . Now take any r -minimal form (8) given by the same sequence. Let $C(G) = \max_{l=1}^{T_r} \frac{\xi_l^{(r)}(G)}{W_r^r(G_0, G)}$. If $\liminf C(G) = 0$, we are done. If not, we have $\liminf C(G) > 0$. It follows that

$$\sum_{l=1}^{T_r} \left(\frac{\xi_l^{(r)}(G)}{C(G)W_r^r(G, G_0)} \right) H_l^{(r)}(x) \rightarrow 0.$$

Moreover, all the coefficients in the above display are bounded from above by 1, one of which is in fact 1. There exists a subsequence of G by which these coefficients have limits, one of which is 1. This is also a contradiction due to the linear independence of functions $H_l^{(r)}(x)$.

(b) Let G be an element in the sequence that admits a r -minimal form such that $\xi_l^{(r)}(G)/W_r^r(G_0, G) \rightarrow 0$ for all $l = 1, \dots, T_r$. It suffices to assume that the basis functions $H_l^{(r)}$ are selected from the collection of partial derivatives of f . We will show that the same sequence of G and the elimination procedure for the r -minimal form can be used to construct a $r - 1$ -minimal form by which

$$\xi_l^{(r-1)}(G)/W_{r-1}^{r-1}(G_0, G) \rightarrow 0$$

for all $l = 1, \dots, T_{r-1}$. There are two possibilities to consider.

First, suppose that each of the r -th partial derivatives of density kernel f (i.e., $\partial^\kappa f / \partial \eta^\kappa$, where $|\kappa| = r$) is not in the linear span of the collection of partial derivatives of f at order $r - 1$ or less. Then, for each $l = 1, \dots, T_{r-1}$, $\xi_l^{(r-1)}(G) = \xi_{l'}^{(r)}(G)$ for some $l' \in [1, T_r]$. Since $W_{r-1}^{r-1}(G, G_0) \gtrsim W_r^r(G, G_0)$, due to the fact that the support points of G and G_0 are in a bounded set, we have that

$$\xi_l^{(r-1)}(G)/W_{r-1}^{r-1}(G_0, G) \lesssim \xi_{l'}^{(r)}(G)/W_r^r(G_0, G)$$

which vanishes by the hypothesis.

Second, suppose that some of the r -th partial derivatives, say, $\partial^{|\kappa|} f / \partial \eta^\kappa$ where $|\kappa| = r$, can be eliminated because they can be represented by a linear combination of a subset of other partial derivatives $H_l^{(r-1)}$ (in addition to possibly a subset of other partial derivatives $H_l^{(r)}$) with corresponding finite coefficients $\alpha_{\kappa, i, l}$. It follows that for each $l = 1, \dots, T_{r-1}$, the coefficient $\xi_l^{(r-1)}(G)$ that defines the $r - 1$ -minimal form is transformed into a coefficient in the r -minimal form by

$$\xi_{l'}^{(r)}(G) := \xi_l^{(r-1)}(G) + \sum_{\kappa; |\kappa|=r} \sum_{i=1}^{k_0} \alpha_{\kappa, i, l} \sum_{j=1}^{s_i} p_{ij} (\Delta \eta_{ij})^\kappa / \kappa!.$$

Since $\xi_{l'}^{(r)}(G)/W_r^r(G, G_0)$ tends to 0, so does $\xi_{l'}^{(r)}(G)/W_{r-1}^{r-1}(G, G_0)$. By Lemma 3.1 for each κ such that $|\kappa| = r$, $\sum_{i=1}^{k_0} \sum_{j=1}^{s_i} p_{ij}(\Delta\eta_{ij})^\kappa / \kappa! = o(D_{r-1}(G_0, G)) = o(W_{r-1}^{r-1}(G, G_0))$. It follows that $\xi_l^{(r-1)}(G)/W_{r-1}^{r-1}(G, G_0)$ tends to 0, for each $l = 1, \dots, T_{r-1}$. This completes the proof. \square

The monotonicity of r -singularity naturally leads to the notion of singularity level of a mixing measure G_0 (and the corresponding parameters) relative to an ambient space \mathcal{G} .

Definition 3.3. *The singularity level of G_0 relative to a given class \mathcal{G} , denoted by $\ell(G_0|\mathcal{G})$, is*

0, if G_0 is not r -singular for any $r \geq 1$;

∞ , if G_0 is r -singular for all $r \geq 1$;

otherwise, the largest natural number $r \geq 1$ for which G_0 is r -singular.

The role of the ambient space \mathcal{G} is critical in determining the singularity level of $G_0 \in \mathcal{G}$. Clearly, if $\mathcal{G} \subset \mathcal{G}'$ are both subsets of probability measures that contain G_0 , r -singularity relative to \mathcal{G} entails r -singularity relative to \mathcal{G}' . This means $\ell(G_0|\mathcal{G}) \leq \ell(G_0|\mathcal{G}')$. Let us look at the following examples.

- Take $\mathcal{G} = \mathcal{E}_{k_0}$, i.e., the setting of e-mixtures. It is easy to verify that if the collection of $\{\partial^\kappa f / \partial \eta^\kappa(x|\eta_j) | j = 1, \dots, k_0; |\kappa| \leq 1\}$ evaluated at G_0 is linearly independent, then G_0 is not 1-singular relative to \mathcal{E}_{k_0} . It follows that $\ell(G_0|\mathcal{G}) = 0$.
- On the other hand, if $\mathcal{G} = \mathcal{O}_k$ for any $k > k_0$, i.e., the setting of o-mixtures. Then it can be shown that G_0 is always 1-singular relative to \mathcal{O}_k for any $k > k_0$. Thus, $\ell(G_0|\mathcal{G}) \geq 1$. Now, if the collection of $\{\partial^\kappa f / \partial \eta^\kappa(x|\eta_j) | j = 1, \dots, k_0; |\kappa| \leq 2\}$ evaluated at G_0 is linearly independent, then it can be observed that G_0 is not 2-singular relative to \mathcal{O}_k . Thus, $\ell(G_0|\mathcal{G}) = 1$.

In fact, the conditions described in the two examples above are referred to as strong identifiability conditions studied by Chen [1995], Nguyen [2013], Ho and Nguyen [2016b]. Our concept of singularity level generalizes such strong identifiability conditions, by allowing us to consider situations where such conditions fail to hold. This is when $\ell(G_0|\mathcal{G}) = 2, 3, \dots, \infty$. The significance of this concept can be appreciated by the following theorem.

Theorem 3.1. *Let \mathcal{G} be a class of probability measures on Θ that have a bounded number of support points, and fix $G_0 \in \mathcal{G}$. Suppose that $\ell(G_0|\mathcal{G}) = r$, for some $0 \leq r \leq \infty$.*

(i) *If $r < \infty$, then $\inf_{G \in \mathcal{G}} \frac{\|p_G - p_{G_0}\|_\infty}{W_s^s(G, G_0)} > 0$ for any $s \geq r + 1$.*

(ii) *If $r < \infty$, then $\inf_{G \in \mathcal{G}} \frac{V(p_G, p_{G_0})}{W_s^s(G, G_0)} > 0$ for any $s \geq r + 1$.*

(iii) *If $1 \leq r < \infty$ and in addition,*

(a) *f is $(r + 1)$ -order differentiable with respect to η and for some constant $c_0 > 0$,*

$$\sup_{\|\eta - \eta'\| \leq c_0} \int_{x \in \mathcal{X}} \left(\frac{\partial^{r+1} f}{\partial \eta^\alpha}(x|\eta) \right)^2 / f(x|\eta') dx < \infty \quad (9)$$

for any $|\alpha| = r + 1$.

(b) There is a sequence $G \in \mathcal{G}$ tending to G_0 in Wasserstein metric W_r and the coefficients of the r -minimal form $\xi_l^{(r)}(G) = 0$ for all $l = 1, \dots, T_r$.

Then, for any $1 \leq s < r + 1$,

$$\liminf_{G \in \mathcal{G}: W_1(G, G_0) \rightarrow 0} \frac{h(p_G, p_{G_0})}{W_1^s(G, G_0)} = 0.$$

(iv) If $r = \infty$ and the conditions (a), (b) in part (iii) hold for any $l \in \mathbb{N}$ (here, the parameter r in these conditions is replaced by l), then the conclusion of part (iii) holds for any $s \geq 1$.

We make a few remarks.

- Part (i) and part (ii) show how the singularity level of G_0 relative to an ambient space \mathcal{G} may be used to translate the convergence of mixture densities (under the sup-norm and the total variation distance) into the convergence of mixing measures under a Wasserstein metric. Part (iii) shows a sufficient condition under which the power $r + 1$ in the bounds from part (i) and (ii) is in fact tight.
- In part (iii) the condition regarding the integrand of the partial derivative of f (cf. Eq. (9)) can be easily checked to be satisfied by many kernels, such as Gaussian kernel, Gamma kernel, Student's kernel, and skewnormal kernel.
- Condition (b) regarding the sequence of G appears somewhat opaque in general, but it will be seen in specific examples for skewnormal mixtures in the sequel. It is sufficient, but not necessary, for verifying the r -singularity of G_0 to construct the sequence of G so that $\xi_l^{(r)}(G) = 0$ for all $1 \leq l \leq T_r$, provided such a sequence exists. This requires finding an appropriate parameterization of a sequence of G tending toward G_0 that satisfy a number of polynomial equations defined in terms of the parameter perturbations.

Proof. Here, we provide the proof for part (i) and (ii) of the theorem. The proof for part (iii) and (iv) is deferred to the Appendix.

(i) It suffices to prove the first inequality for $s = r + 1$. Firstly, we will demonstrate that

$$\liminf_{G \in \mathcal{G}: W_s(G, G_0) \rightarrow 0} \|p_G - p_{G_0}\|_\infty / W_s^s(G, G_0) > 0.$$

If this is not true, then there exists a sequence of G such that $W_s(G, G_0) \rightarrow 0$, and for any x , $(p_G(x) - p_{G_0}(x)) / W_s^s(G, G_0) \rightarrow 0$. Take any s -minimal form for this ratio, we have

$$\frac{p_G(x) - p_{G_0}(x)}{W_s^s(G, G_0)} = \sum_{l=1}^{T_s} \left(\frac{\xi_l^{(s)}(G)}{W_s^s(G, G_0)} \right) H_l^{(s)}(x) + o(1) \rightarrow 0.$$

For each G in the sequence, let $C(G) = \max_l \frac{\xi_l^{(s)}(G)}{W_s^s(G, G_0)}$. If $\liminf C(G) = 0$, then this means G_0 is s -singular, so $\ell(G_0 | \mathcal{G}) \geq s$. This violates the given assumption. So we have $\liminf C(G) > 0$. It follows that

$$\sum_{l=1}^{T_s} \left(\frac{\xi_l^{(s)}(G)}{C(G) W_s^s(G, G_0)} \right) H_l^{(s)}(x) \rightarrow 0.$$

Moreover, all coefficients in the above display are bounded from above by 1, one of which is in fact 1. There exists a subsequence of G by which these coefficients have a limit, one of which is 1. This is also a contradiction due to the linear independence of functions $H_l^{(s)}$.

Therefore, we can find a positive number ϵ_0 such that $\|p_G - p_{G_0}\|_\infty \gtrsim W_s^s(G, G_0)$ for any $W_s(G, G_0) \leq \epsilon_0$. Now, to obtain the conclusion of part (i), it suffices to demonstrate that

$$\inf_{G \in \mathcal{G}: W_s(G, G_0) > \epsilon_0} \|p_G - p_{G_0}\|_\infty / W_s^s(G, G_0) > 0.$$

If this is not the case, there is a sequence G' such that $W_s(G', G_0) > \epsilon_0$ and $\|p_{G'} - p_{G_0}\|_\infty / W_s^s(G', G_0) \rightarrow 0$. Since Θ is compact and \mathcal{G} contains only probability measures with bounded number of support points in Θ , we can find $G^* \in \mathcal{G}$ such that $W_s(G', G^*) \rightarrow 0$ and $W_s(G^*, G_0) \geq \epsilon_0$. As $W_s(G', G_0) \rightarrow W_s(G^*, G_0) > 0$, we have $\|p_{G'} - p_{G_0}\|_\infty \rightarrow 0$. Now, due to the first order uniform Lipschitz condition of f , we obtain $p_{G'}(x) \rightarrow p_{G^*}(x)$ for all $x \in \mathcal{X}$. Thus, $p_{G^*}(x) = p_{G_0}(x)$ for almost all $x \in \mathcal{X}$, which entails that $G^* = G_0$, a contradiction. This completes the proof.

(ii) Turning to the second inequality, we also firstly demonstrate that

$$\liminf_{G \in \mathcal{G}: W_s(G, G_0) \rightarrow 0} V(p_G, p_{G_0}) / W_s^s(G, G_0) > 0.$$

If it is not true, then we have a sequence of G such that $W_s(G, G_0) \rightarrow 0$ and $V(p_G, p_{G_0}) / W_s^s(G, G_0) \rightarrow 0$. By Fatou's lemma

$$0 = \liminf \frac{V(p_G, p_{G_0})}{C(G)W_s^s(G, G_0)} \geq \int \liminf_G \left| \frac{\xi_l^{(s)}(G)}{C(G)W_s^s(G, G_0)} H_l^{(s)}(x) \right| dx.$$

The integrand must be zero for almost all x , leading to a contradiction as before. Hence, to obtain the conclusion of part (ii), we only need to show that

$$\inf_{G \in \mathcal{G}: W_s(G, G_0) > \epsilon_0} V(p_G, p_{G_0}) / W_s^s(G, G_0) > 0.$$

where $\epsilon_0 > 0$ such that $V(p_G, p_{G_0}) \gtrsim W_s^s(G, G_0)$ for any $W_s(G, G_0) \leq \epsilon_0$. If it is not true, by using the same argument as that of part (i), there is a sequence of G' such that $W_s(G', G^*) \rightarrow 0$, $V(p_{G'}, p_{G_0}) \rightarrow 0$, while $W_s(G^*, G_0) \geq \epsilon_0$ and $p_{G'}(x) \rightarrow p_{G^*}(x)$ for all $x \in \mathcal{X}$. By Fatou's lemma,

$$0 = \liminf V(p_{G'}, p_{G_0}) \geq \int \liminf |p_{G'}(x) - p_{G_0}(x)| dx = V(p_{G^*}, p_{G_0}),$$

which leads to $G^* = G_0$, a contradiction. We obtain the conclusion of this part. \square

We are ready to state the impact of the singularity level of mixing measure G_0 relative to an ambient space \mathcal{G} on the rate of convergence for an estimate of G_0 , where $\mathcal{G} = \mathcal{E}_{k_0}$ in e-mixtures, and $\mathcal{G} = \mathcal{O}_k$ in o-mixtures. Let \mathcal{G} be structured into a sieve of subsets defined by the maximum singularity level relative to \mathcal{G} .

$$\mathcal{G} = \bigcup_{r=1}^{\infty} \mathcal{G}_r, \text{ where } \mathcal{G}_r := \left\{ G \in \mathcal{G} \mid \ell(G|\mathcal{G}) \leq r \right\}, r = 0, 1, \dots, \infty.$$

The first part of the following theorem gives a minimax lower bound for the estimation of the mixing measure G_0 , given that the singularity level of G_0 is known up to a constant $r \geq 1$. The second part gives a quick result on the convergence rate of a point estimate such as the MLE.

Theorem 3.2. (a) Fix $r \geq 1$. Assume that for any $G_0 \in \mathcal{G}_r$, the conclusion of part (iii) of Theorem 3.1 holds for \mathcal{G}_r (i.e., \mathcal{G} is replaced by \mathcal{G}_r in that theorem). Then, for any $s \in [1, r+1)$ there holds

$$\inf_{\hat{G}_n \in \mathcal{G}_r} \sup_{G_0 \in \mathcal{G}_r} E_{p_{G_0}} W_s(\hat{G}_n, G_0) \gtrsim n^{-1/2s}.$$

Here, the infimum is taken over all sequences of estimates $\hat{G}_n \in \mathcal{G}_r$ and $E_{p_{G_0}}$ denotes the expectation taken with respect to product measure with mixture density $p_{G_0}^n$.

(b) Let $G_0 \in \mathcal{G}_r$ for some fixed $r \geq 1$. Let $\hat{G}_n \in \mathcal{G}_r$ be a point estimate for G_0 , which is obtained from an n -sample of i.i.d. observations drawn from p_{G_0} . As long as $h(p_{\hat{G}_n}, p_{G_0}) = O_P(n^{-1/2})$, we have

$$W_{r+1}(\hat{G}_n, G_0) = O_P(n^{-1/2(r+1)}).$$

Proof. Part (a) of this theorem is a consequence of the conclusion of Theorem 3.1, part (iii). The proof of this fact is quite standard, and similar to that of Theorem 1.1. of [Ho and Nguyen, 2016a] and is omitted. Part (b) follows immediately from part (ii) of Theorem 3.1, as we have $h(p_{\hat{G}_n}, p_{G_0}) \geq V(p_{\hat{G}_n}, p_{G_0}) \gtrsim W_{r+1}^{r+1}(\hat{G}_n, G_0)$. \square

We conclude this section with some comments. It is well-known that many density estimation methods, such as MLE and Bayesian estimation applied to a compact parameter space for parametric mixture models, guarantee a root- n rate (up to a logarithmic term) of convergence under Hellinger distance metric on the density functions (cf. [van de Geer, 2000, Ghosal and van der Vaart, 2001, DasGupta, 2008]). It follows that, as far as we are concerned, the remaining work in establishing the convergence behavior of parameter estimation (as opposed to density estimation) lies in the calculation of the singularity levels, i.e., the identification of sets \mathcal{G}_r . For skewnormal mixtures, such calculations will be carried out in Section 4 and Section 5. For the settings of G_0 where we are able to obtain the exact singularity levels, we can also construct the sequence of G required by part (iii) of Theorem 3.1. Whenever the exact singularity level is obtained, we automatically obtain a minimax lower bound and a matching upper bound for MLE convergence rate under a Wasserstein distance metric, thanks to the above theorem. In some cases, however, the singularity level of G_0 may be not determined exactly, but only an upper bound is given. In such cases, only an upper bound to the convergence rate of the MLE can be obtained, while minimax lower bounds may be unknown.

3.3 Construction of r -minimal forms

As we mentioned above, a simple way of constructing an r -minimal form is to select a subset of partial derivatives of f taken up to order r such that all these functions are linearly independent. A simple procedure is to start from the smallest order $r = 1$ and then move up to $r = 2, 3, \dots$ and so on. For each r , assume that we have obtained a linearly independent subset of partial derivatives up to order $r - 1$. Now, going over the ordered list of r -th partial derivatives: $\{\partial^{|\kappa|} f / \partial \eta^\kappa | \kappa \in \mathbb{N}^d, |\kappa| = r\}$. For each κ such that $|\kappa| = r$, if the partial derivative of f of order κ can be expressed as a linear combination of other partial derivatives already selected, then this one is eliminated. The process goes on until we exhaust the list of the partial derivatives.

Example 3.2. Continuing from Example 3.1, suppose that G_0 satisfies Eq. (1). According to the proof of Lemma 4.1, we can choose $\alpha_{4k} \neq 0$, so the partial derivative may be eliminated via the reduction:

$$\frac{\partial f(x|\eta_k^0)}{\partial m} = - \sum_{j=1}^k \frac{\alpha_{1j}}{\alpha_{4k}} f(x|\eta_j^0) + \frac{\alpha_{2j}}{\alpha_{4k}} \frac{\partial f(x|\eta_j^0)}{\partial \theta} + \frac{\alpha_{3j}}{\alpha_{4k}} \frac{\partial f(x|\eta_j^0)}{\partial v} - \sum_{j=1}^{k-1} \frac{\alpha_{4j}}{\alpha_{4k}} \frac{\partial f(x|\eta_j^0)}{\partial m}$$

Note that this elimination step is valid only for a subset of $G_0 = G(\mathbf{p}^0, \boldsymbol{\eta}^0)$ for which Eq. (1) holds. That is, only if $P_1(\boldsymbol{\eta}^0) = 0$ or $P_2(\boldsymbol{\eta}^0) = 0$.

Example 3.3. If $f(x|\eta) = f(x|\theta, v, m)$ where $m = 0$, the skewnormal kernel becomes the Gaussian kernel. Due to (3), all partial derivatives with respect to both θ and v can be eliminated via the following reduction: for any $\kappa_1, \kappa_2 \in \mathbb{N}$, for any $j = 1, \dots, k_0$,

$$\frac{\partial^{\kappa_1+\kappa_2} f(x|\eta_j^0)}{\partial \theta^{\kappa_1} \partial v^{\kappa_2}} = \frac{1}{2^{\kappa_2}} \frac{\partial^{\kappa_1+2\kappa_2} f(x|\eta_j^0)}{\partial \theta^{\kappa_1+2\kappa_2}}.$$

Thus, this elimination is valid for all parameter values $(\mathbf{p}^0, \boldsymbol{\eta}^0)$, and r -minimal forms for all orders.

Example 3.4. For the skewnormal kernel density $f(x|\eta) = f(x|\theta, v, m)$, Eq. (2) yields the following reductions: for any $j = 1, \dots, k_0$, any $\eta = (\theta, v, m) = \eta_j^0 = (\theta_j^0, v_j^0, m_j^0)$ such that $m \neq 0$

$$\frac{\partial^2 f}{\partial \theta^2} = 2 \frac{\partial f}{\partial v} - \frac{m^3 + m}{v} \frac{\partial f}{\partial m}, \quad (10)$$

$$\frac{\partial^2 f}{\partial v \partial m} = -\frac{1}{v} \frac{\partial f}{\partial m} - \frac{m^2 + 1}{2vm} \frac{\partial^2 f}{\partial m^2}. \quad (11)$$

This results in a ripple effect on subsequent eliminations at higher orders. For examples, partial derivatives up to the third order of f evaluated at $\eta = \eta_j^0 = (\theta_j^0, v_j^0, m_j^0)$ for any $j = 1, \dots, k_0$ where $m_j^0 \neq 0$ can be expressed as follows:

$$\begin{aligned} \frac{\partial^3 f}{\partial \theta^3} &= 2 \frac{\partial^2 f}{\partial \theta \partial v} - \frac{m^3 + m}{v} \frac{\partial^2 f}{\partial \theta \partial m}, \\ \frac{\partial^3 f}{\partial \theta^2 \partial v} &= 2 \frac{\partial^2 f}{\partial v^2} + \frac{m^3 + m}{v^2} \frac{\partial f}{\partial m} - \frac{m^3 + m}{v} \frac{\partial^2 f}{\partial v \partial m}, \\ \frac{\partial^3 f}{\partial \theta^2 \partial m} &= 2 \frac{\partial^2 f}{\partial v \partial m} - \frac{3m^2 + 1}{v} \frac{\partial f}{\partial m} - \frac{m^3 + m}{v} \frac{\partial^2 f}{\partial m^2}, \\ \frac{\partial^3 f}{\partial v \partial m^2} &= -\frac{m^2 + 1}{2vm} \frac{\partial^3 f}{\partial m^3} - \frac{3m^2 - 1}{2vm^2} \frac{\partial^2 f}{\partial m^2}, \\ \frac{\partial^3 f}{\partial v^2 \partial m} &= -\frac{2}{v} \frac{\partial^2 f}{\partial v \partial m} - \frac{m^2 + 1}{2vm} \frac{\partial^3 f}{\partial v \partial m^2} \\ &= \frac{(m^2 + 1)^2}{4v^2 m^2} \frac{\partial^3 f}{\partial m^3} + \frac{(m^2 + 1)(7m^2 - 1)}{4m^3 v^2} \frac{\partial^2 f}{\partial m^2} + \frac{2}{v^2} \frac{\partial f}{\partial m}, \\ \frac{\partial^3 f}{\partial \theta \partial v \partial m} &= -\frac{m^2 + 1}{2vm} \frac{\partial^3 f}{\partial \theta \partial m^2} - \frac{1}{v} \frac{\partial^2 f}{\partial \theta \partial m}. \end{aligned} \quad (12)$$

All three examples above demonstrate how the dependence among partial derivatives of kernel density f , among different orders κ , and among those evaluated at different component i , has a deep impact on the representation of r -minimal forms.

In general, the r -minimal form (8) may be expressed somewhat more explicitly as follows

$$\frac{p_G(x) - p_{G_0}(x)}{W_r^r(G, G_0)} = \sum_{(i, \kappa) \in \mathcal{I}, \mathcal{K}} \frac{\xi_{i, \kappa}^{(r)}(G)}{W_r^r(G_0, G)} H_{i, \kappa}^{(r)}(x|G_0) + \sum_{i=1}^{k_0} \frac{\zeta_i^{(r)}(G)}{W_r^r(G_0, G)} f(x|\eta_i^0) + o(1).$$

where $\mathcal{I} \subset \{1, \dots, k_0\}$ and $\mathcal{K} \subset \mathbb{N}^d$ of elements κ such that $|\kappa| \leq r$. It is emphasized that the sets \mathcal{I} and \mathcal{K} are specific to a particular r -minimal form under consideration. $H_{i, \kappa}^{(r)}$ are a collection of linearly independent partial derivatives of f that are also independent of all functions $f(x|\eta_i^0)$. $H_{i, \kappa}^{(r)}$ are taken

from the collection of partial derivatives with order at most r . We also observe that $\xi_{i,\kappa}^{(r)}$ and $\zeta_i^{(r)}$ take the following polynomial forms:

$$\xi_{i,\kappa}^{(r)}(G) = \sum_{j=1}^{s_i} \frac{p_{ij}(\Delta\eta_{ij})^\kappa}{\kappa!} + \sum_{i',\kappa'} \beta_{i,\kappa,i',\kappa'}(G_0) \sum_{j=1}^{s_{i'}} \frac{p_{ij}(\Delta\eta_{ij})^{\kappa'}}{\kappa'!}, \quad (13)$$

$$\zeta_i^{(r)}(G) = \Delta p_i + \sum_{i',\kappa'} \gamma_{i,\kappa,i',\kappa'}(G_0) \sum_{j=1}^{s_{i'}} \frac{p_{ij}(\Delta\eta_{ij})^{\kappa'}}{\kappa'!}. \quad (14)$$

In the right hand side of each of the last two equations, i' is taken from a subset of $\{1, \dots, k_0\}$ and κ' is from a subset of \mathbb{N}^d such that $|\kappa| \leq |\kappa'| \leq r$. The actual detail of these subsets depend on the specific elimination scheme leading to the r -minimal form. Likewise, the non-zero coefficients $\beta_{i,\kappa,i',\kappa'}(G_0)$ and $\gamma_{i,\kappa,i',\kappa'}(G_0)$ arise from the specific elimination scheme. We include argument G_0 in these coefficients to highlight the fact that they may be dependent on the atoms of G_0 (cf. Example 3.2 and 3.4).

By the definition of r -singularity for any $r \geq 1$, G_0 is r -singular relative to \mathcal{G} if there exists a sequence of G tending to G_0 in the ambient space \mathcal{G} such that the sequences of semipolynomial fractions, namely, $\xi_{i,\kappa}^{(r)}(G)/W_r^r(G, G_0)$ and $\zeta_i^{(r)}(G)/W_r^r(G, G_0)$ (whose numerators are given by Eq. (13) and Eq. (14)), must vanish. As a consequence, the question of r -singularity for a given element G_0 is determined by the limiting behavior of a finite collection of infinite sequences of semipolynomial fractions.

3.4 Polynomial limits of r -minimal form coefficients

It is worth noting that the limiting behavior of semipolynomial fractions described above is independent of a particular choice of the r -minimal form, in a sense that we now explain. In part (a) of Lemma 3.2, we established an invariance property of the r -singularity, which does not depend on a specific form of the r -minimal form. Let us restrict the basis functions to be members of the collection of all partial derivatives of f up to order r . In the proof of part (b) of Lemma 3.2 it was shown that the coefficients $\xi_l^{(r)}(G)$ have to be elements of a set of polynomials of $\Delta\eta_{ij}$, Δp_i , and p_{ij} , which are closed under linear combinations of its elements. Let us denote this set by $\mathcal{P}(G, G_0)$, which is invariant with respect to any specific choice of the basis functions (from the collection of partial derivatives) for the r -minimal form. Moreover, G_0 is r -singular if and only if a sequence of G tending to G_0 in W_r can be constructed such that for any element $\xi_l^{(r)}(G) \in \mathcal{P}(G, G_0)$, we have $\xi_l^{(r)}(G)/W_r^r(G, G_0) \rightarrow 0$. Equivalently,

$$\xi_l^{(r)}(G)/D_r(G, G_0) \rightarrow 0 \text{ for all } \xi_l^{(r)}(G) \in \mathcal{P}(G, G_0). \quad (15)$$

Extracting the limits of a single multivariate semipolynomial fraction (a.k.a. rational semipolynomial functions) is quite challenging in general, due to the interaction among multiple variables involved [Xiao and Zeng, 2014]. Analyzing the limits of not one but a collection of multivariate rational semipolynomials is considerably more difficult. To obtain meaningful and concrete results, we need to consider specific systems of multivariate rational semipolynomials that arise from the r -minimal form.

In the remainder of this paper we will proceed to do just that. By working with a specific choice of kernel density f (the skewnormal), it will be shown that under the compactness of the parameter spaces, one can extract a subset of limits from the system of rational semipolynomials $\xi_l^{(r)}(G)/D_r(G, G_0)$. These limits are expressed as a system of polynomials admitting non-trivial solutions. For a given $r \geq 1$, if the extracted system of polynomial limits does not contain admissible solutions, then it means that there does not exist any sequence of mixing measures G for which a valid r -minimal form can be

constructed, because (15) is not fulfilled. This would entail the upper bound $\ell(G_0|\mathcal{G}) < r$. On the other hand, if the extracted system of polynomial limits does contain at least one admissible solution, this is a hint that the r -singularity level of G_0 relative to the ambient space G *might* hold. Whether this is actually the case or not requires an explicit construction of a sequence of $G \in \mathcal{G}$ (often building upon the admissible solutions of the polynomial limits) and then the verification that condition (15) indeed holds. For the verification purpose, it is sufficient (and simpler) to work with a specific choice of r -minimal form, as Definition 3.2 allows.

The foregoing description, along with the presentation in the previous subsection on the construction of r -minimal forms, provides the outline of a general procedure which links the determination of the singularity level to the solvability of a system of polynomial limits. This procedure will be illustrated in the next sections for the remarkable world of mixtures of skewnormal distributions.

4 O-mixtures of skewnormal distributions

In this section, we focus on the o-mixture setting of skewnormal distributions. To avoid a heavy dose of technicality, we study the singularity level of $G_0 \in \mathcal{E}_{k_0}$ relative to ambient space \mathcal{O}_{k,c_0} for some $k > k_0$ and small constant $c_0 > 0$, where $\mathcal{O}_{k,c_0} \subset \mathcal{O}_k$ contains only (discrete) probability measures whose point masses are bounded from below by c_0 . Moreover, we will analyze the singularity level of $G_0 \in \mathcal{S}_0$, a subset to be defined shortly by (16). This case is interesting in that it illustrates the full power of the general method of analysis that was described in Section 3 in a concrete fashion. Due to the complex nature and space constraints, we will not report any result on the case where G_0 is in the complement of \mathcal{S}_0 .² Instead, in Section 5 we study the singularity level of G_0 relative to the smaller ambient space \mathcal{E}_{k_0} (that is, e-mixture setting), for which a more complete picture of the singularity structure is achieved.

Lemma 4.1. *For skewnormal density kernel $f(x|\boldsymbol{\eta})$, the collection of $\{\partial^\kappa f / \partial \eta^\kappa(x|\eta_j) | j = 1, \dots, k_0; 0 \leq |\kappa| \leq 1\}$ is not linearly independent if and only if $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)$ are the zeros of either polynomial P_1 or P_2 , which are defined as follows:*

$$\text{Type A: } P_1(\boldsymbol{\eta}) = \prod_{j=1}^{k_0} m_j.$$

$$\text{Type B: } P_2(\boldsymbol{\eta}) = \prod_{1 \leq i \neq j \leq k_0} \left\{ (\theta_i - \theta_j)^2 + \left[\sigma_i^2(1 + m_j^2) - \sigma_j^2(1 + m_i^2) \right]^2 \right\}.$$

This lemma leads us to consider

$$\mathcal{S}_0 = \left\{ G = G(\mathbf{p}, \boldsymbol{\eta}) \mid (\mathbf{p}, \boldsymbol{\eta}) \in \Omega, P_1(\boldsymbol{\eta}) \neq 0, P_2(\boldsymbol{\eta}) \neq 0 \right\}. \quad (16)$$

In the o-mixture setting, we will see that $\ell(G_0|\mathcal{O}_{k,c_0})$ may grow with $k - k_0$, the number of extra mixing components. The main exercise is to arrive at a suitable r -minimal form, for which the vanishing behavior of its coefficients can be analyzed. Section 3.3 describes a general strategy for the construction of r -minimal form based on the partial derivatives of the density kernel f with respect to the parameters $\boldsymbol{\eta} = (\theta, v, m)$ up to order r .

For skewnormal kernel density f , the following lemma provides an explicit form for reducing a partial derivative of f to other partial derivatives of lower orders.

²Interested readers may consult Section 6 in technical report [Ho and Nguyen, 2016c] for such results.

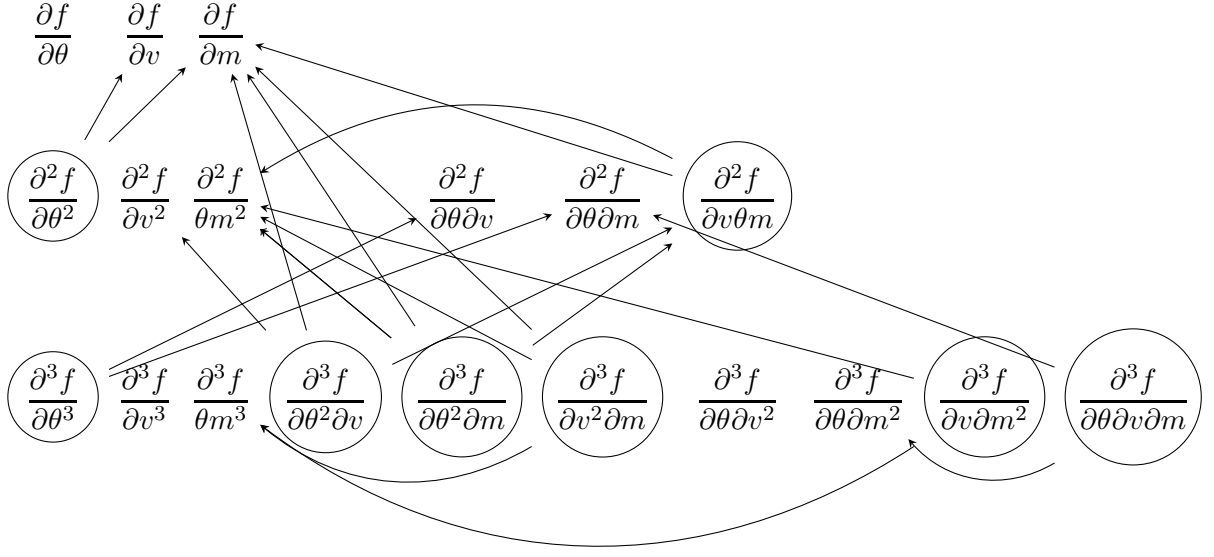


Figure 1: The illustration of the elimination steps from a complete collection of derivatives of f up to the order 3 to a reduced system of linearly independent partial derivatives, cf. Lemma 4.3. The circled derivatives are eliminated from the partial derivatives present in the 3-minimal form. $A \rightarrow B$ means that B is involved in the representation of A under the reduction.

Lemma 4.2. For any $r \geq 1$, denote

$$\begin{aligned} A_1^r &= \{(\alpha_1, \alpha_2, \alpha_3) : \alpha_1 \leq 1, \alpha_3 = 0, \text{ and } |\alpha| \leq r\}. \\ A_2^r &= \{(\alpha_1, \alpha_2, \alpha_3) : \alpha_1 \leq 1, \alpha_2 = 0, \alpha_3 \geq 1, \text{ and } |\alpha| \leq r\}. \\ \mathcal{F}_r &= A_1^r \cup A_2^r. \end{aligned}$$

Let $f(x|\eta) = f(x|\theta, v, m)$ denote the skewnormal kernel. Then, for any $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{N}^3$ and $m \neq 0$, there holds

$$\frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} \partial v^{\alpha_2} \partial m^{\alpha_3}} = \sum_{\kappa \in \mathcal{F}_{|\alpha|}} \frac{P_{\alpha_1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2} \partial m^{\kappa_3}},$$

where, $P_{\alpha_1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)$, $H_{\alpha_1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)$, and $Q_{\alpha_1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)$ are polynomials in terms of m, v respectively.

Next, we show that the partial derivatives on the RHS of the above identity are in fact linearly independent, under additional assumptions on G_0 .

Lemma 4.3. Recall the notation from Lemma 4.2. If $G_0 \in \mathcal{S}_0$, then for any $r \geq 1$, the collection of partial derivatives of the skewnormal density kernel $f(x|\eta)$, namely

$$\left\{ \frac{\partial^{|\kappa|} f(x|\eta)}{\theta^{\kappa_1} v^{\kappa_2} m^{\kappa_3}} \mid \kappa = (\kappa_1, \kappa_2, \kappa_3) \in \mathcal{F}_r, \eta = \eta_1^0, \dots, \eta_{k_0}^0 \right\}$$

is linearly independent.

Figure 1 gives an illustration of Lemma 4.3 when $r = 3$. Armed with the foregoing lemmas we can easily obtain a suitable r -minimal form for the mixture densities of skewnormals.

4.1 Special cases

To illustrate our techniques and results, consider a special case in which G_0 has exactly one atom, and $k = k_0 + 1 = 2$. The general results will be presented in Section 4.2.

G_0 is 1-singular $G_0 \in \mathcal{S}_0$ implies that all first order derivatives of f are linearly independent. Hence, from Eq. (8), the 1-minimal form takes the form:

$$\begin{aligned} \frac{p_G(x) - p_{G_0}(x)}{W_1(G, G_0)} &\asymp \frac{1}{W_1(G, G_0)} \left(\Delta p_1 \cdot f(x|\eta_1^0) + \sum_{i=1}^2 p_{1i} \Delta \theta_{1i} \frac{\partial f}{\partial \theta}(x|\eta_1^0) \right. \\ &\quad \left. + \sum_{i=1}^2 p_{1i} \Delta v_{1i} \frac{\partial f}{\partial v}(x|\eta_1^0) + \sum_{i=1}^2 p_{1i} \Delta m_{1i} \frac{\partial f}{\partial m}(x|\eta_1^0) \right) + o(1). \end{aligned} \quad (17)$$

Since $k = 2$ and $k_0 = 1$, we have $\Delta p_1 = 0$. A sequence of G can be chosen so that $\sum_{i=1}^2 p_{1i} \Delta \theta_{1i} = 0$, $\sum_{i=1}^2 p_{1i} \Delta v_{1i} = 0$, $\sum_{i=1}^2 p_{1i} \Delta m_{1i} = 0$. Clearly, all of the coefficients in (8) are 0. Hence G_0 is 1-singular relative to \mathcal{O}_{2, c_0} .

G_0 is 2-singular Using the method of elimination described in Example 3.4 we obtain the following 2-minimal form:

$$\frac{1}{W_2^2(G, G_0)} \left(\sum_{\kappa \in \mathcal{F}_2} \xi_{\kappa_1, \kappa_2, \kappa_3}^{(2)} \frac{\partial^{|\alpha|} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2} \partial m^{\kappa_3}}(x|\eta_1^0) \right) + o(1), \quad (18)$$

where $\xi_{\kappa_1, \kappa_2, \kappa_3}^{(2)}$ are given by

$$\begin{aligned} \xi_{1,0,0}^{(2)} &= \sum_{i=1}^2 p_{1i} \Delta \theta_{1i}, \quad \xi_{0,1,0}^{(2)} = \sum_{i=1}^2 p_{1i} \Delta v_{1i} + \sum_{i=1}^2 p_{1i} (\Delta \theta_{1i})^2, \\ \xi_{0,0,1}^{(2)} &= -\frac{(m_1^0)^3 + m_1^0}{2v_1^0} \sum_{i=1}^2 p_{1i} (\Delta \theta_{1i})^2 - \frac{1}{v_1^0} \sum_{i=1}^2 p_{1i} \Delta v_{1i} \Delta m_{1i} + \sum_{i=1}^2 p_{1i} \Delta m_{1i}, \\ \xi_{0,2,0}^{(2)} &= \sum_{i=1}^2 p_{1i} (\Delta v_{1i})^2, \quad \xi_{0,0,2}^{(2)} = -\frac{(m_1^0)^2 + 1}{2v_1^0 m_1^0} \sum_{i=1}^2 p_{1i} \Delta v_{1i} \Delta m_{1i} + \sum_{i=1}^2 p_{1i} \Delta (m_{1i})^2, \\ \xi_{1,1,0}^{(2)} &= \sum_{i=1}^2 p_{1i} \Delta \theta_{1i} \Delta v_{1i}, \quad \xi_{1,0,1}^{(2)} = \sum_{i=1}^2 p_{1i} \Delta \theta_{1i} \Delta m_{1i}. \end{aligned}$$

Note in particular the formulas for $\xi_{0,1,0}^{(2)}$, $\xi_{0,0,1}^{(2)}$ and $\xi_{0,0,2}^{(2)}$ are the results of reduction equation (10). It remains to construct a sequence of G tending to G_0 so that $\xi_{\kappa}^{(2)}/W_2^2(G, G_0)$ vanish for all $\kappa = (\kappa_1, \kappa_2, \kappa_3) \in \mathcal{F}_2$. Define

$$\overline{M} = \max \left\{ |\Delta \theta_{11}|, |\Delta \theta_{12}|, |\Delta v_{11}|^{1/2}, |\Delta v_{12}|^{1/2}, |\Delta m_{11}|^{1/2}, |\Delta m_{12}|^{1/2} \right\}.$$

Then, it can be observed that $W_2^2(G, G_0) \gtrsim \overline{M}^2$ and $\xi_{\kappa_1, \kappa_2, \kappa_3}^{(2)} = O(\overline{M}^{\kappa_1 + 2\kappa_2 + 2\kappa_3})$. So, for any $\kappa \in \mathcal{F}_2$ such that $\kappa_1 + 2\kappa_2 + 2\kappa_3 \geq 3$, as $\xi_{\kappa_1, \kappa_2, \kappa_3}^{(2)} = O(\overline{M}^s)$ where $s \geq 3$, it implies that

$\xi_{\kappa_1, \kappa_2, \kappa_3}^{(2)}/W_2^2(G, G_0) \rightarrow 0$. So we only need to consider the coefficients where $\kappa_1 + 2\kappa_2 + 2\kappa_3 \leq 2$ and $\kappa_1 \leq 1$. They are $\xi_{1,0,0}^{(2)}/W_2^2(G, G_0)$, $\xi_{0,1,0}^{(2)}/W_2^2(G, G_0)$, and $\xi_{0,0,1}^{(2)}/W_2^2(G, G_0)$. Now, by dividing both the numerator and denominator of each of these coefficients by \overline{M} , \overline{M}^2 , and \overline{M}^2 , respectively, we extract the following system of polynomial limits:

$$\begin{aligned} d_1^2 a_1 + d_2^2 a_2 &= 0, \\ d_1^2 a_1^2 + d_2^2 a_2^2 + d_1^2 b_1 + d_2^2 b_2 &= 0, \\ -\frac{(m_1^0)^3 + m_1^0}{2v_1^0}(d_1^2 a_1^2 + d_2^2 a_2^2) + d_1^2 c_1 + d_2^2 c_2 &= 0, \end{aligned} \quad (19)$$

where $\Delta\theta_{1i}/\overline{M} \rightarrow a_i$, $\Delta v_{1i}/\overline{M}^2 \rightarrow b_i$, $\Delta m_{1i}/\overline{M}^2 \rightarrow c_i$, $p_{1i} \rightarrow d_i^2$ for all $1 \leq i \leq 2$. One solution to the above system of polynomial equations is $d_1 = d_2$, $a_1 = -a_2$, $b_1 = b_2 = a_1^2/2$, $c_1 = c_2 = -(m_1^0)^3 + m_1^0/2v_1^0$. It follows that if we choose the sequence of G so that $p_{11} = p_{12} = 1/2$, $\Delta\theta_{11} = -\Delta\theta_{12}$, $\Delta v_{11} = \Delta v_{12} = (\Delta\theta_{11})^2/2$, and $\Delta m_{11} = \Delta m_{12} = (\Delta\theta_{11})^2(-(m_1^0)^3 + m_1^0)/2v_1^0$, then all coefficients of the 2-minimal form vanish. Hence, G_0 is 2-singular relative to \mathcal{O}_{2,c_0} .

G_0 is 3-singular The proof for this is similar. A 3-minimal form can be obtained by applying the reductions (12), which eliminate all third order partial derivatives in terms of lower order ones that are in fact linearly independent by the condition that $G_0 \in \mathcal{S}_0$. As in the foregoing paragraphs, we can obtain a system of polynomials that turn out to share the same solution as the one described. This leads to the same choice of sequence for G according to which all coefficients of the 3-minimal form vanish. Thus, G_0 is 3-singular relatively to \mathcal{O}_{k,c_0} .

G_0 is not 4-singular Applying the same approach to obtain a 4-minimal form and their rational semipolynomial coefficients, from which we extract the following system of real polynomial limits:

$$\begin{aligned} d_1^2 a_1 + d_2^2 a_2 &= 0, \\ d_1^2 a_1^2 + d_2^2 a_2^2 + d_1^2 b_1 + d_2^2 b_2 &= 0, \\ -\frac{(m_1^0)^3 + m_1^0}{2v_1^0}(d_1^2 a_1^2 + d_2^2 a_2^2) + d_1^2 c_1 + d_2^2 c_2 &= 0, \\ \frac{1}{3}(d_1^2 a_1^3 + d_2^2 a_2^3) + d_1^2 a_1 b_1 + d_2^2 a_2 b_2 &= 0, \\ -\frac{(m_1^0)^3 + m_1^0}{6v_1^0}(d_1^2 a_1^3 + d_2^2 a_2^3) + d_1^2 a_1 c_1 + d_2^2 a_2 c_2 &= 0, \\ \frac{1}{6}(d_1^2 a_1^4 + d_2^2 a_2^4) + d_1^2 a_1^2 b_1 + d_2^2 a_2^2 b_2 + \frac{1}{2}(d_1^2 b_1^2 + d_2^2 b_2^2) &= 0, \\ \frac{((m_1^0)^3 + m_1^0)^2}{12(v_1^0)^2}(d_1^2 a_1^4 + d_2^2 a_2^4) - \frac{(m_1^0)^3 + m_1^0}{v_1^0}(d_1^2 a_1^2 c_1 + d_2^2 a_2^2 c_2) - \\ \frac{(m_1^0)^2 + 1}{v_1^0 m_1^0}(d_1^2 b_1 c_1 + d_2^2 b_2 c_2) + d_1^2 c_1^2 + d_2^2 c_2^2 &= 0, \end{aligned} \quad (20)$$

such that at least one among $a_1, a_2, b_1, b_2, c_1, c_2$ is non-zero and $d_1, d_2 \neq 0$. At the first glance, the behavior of this system may be dependent on the specific value of v_1^0, m_1^0 . However, if we remove the third, fifth and eighth equations, we obtain a system of real polynomials that does not depend on the specific value of G_0 . In fact, it can be verified that this system does not admit any non-trivial real solution. Thus, there does not exist any sequence of $G \in \mathcal{O}_{2,c_0}$ according to which all coefficients of the 4-minimal form vanish. So, G_0 is *not* 4-singular. We conclude that $\ell(G_0|\mathcal{O}_{2,c_0}) = 3$.

We end this illustrative exercise with a remark. The fact that there exists a subset of the limiting polynomials of the coefficients of r -minimal forms that do not depend on specific value of G_0 is very useful, because it allows us to provide an upper bound on the singularity level the holds uniformly for all $G_0 \in \mathcal{S}_0$. It is interesting to note that this subset of polynomials also arises from the same analysis applied to the Gaussian kernels studied by Ho and Nguyen [2016a]. This observation can be partially explained by the fact that Gaussian kernels are a special case of skewnormal kernels with zero skewness. A highly nontrivial consequence from this observation is that the singularity level in a skewnormal mixture is always bounded from above that the singularity level in a Gaussian mixture. Thanks to Theorem 3.2 we arrive at a somewhat surprising conclusion that the MLE and minimax bounds for parameter estimation in skewnormal o-mixtures are generally *faster* than that of Gaussian o-mixtures. Now we are ready for results for the general setting of $G_0 \in \mathcal{S}_0$, which also articulates this remark more precisely.

4.2 General results

In this section we shall present results on $\ell(G_0|\mathcal{O}_{k,c_0})$ for the general case $k > k_0$. To do so, we shall define the system of the limiting polynomials that characterizes the singularity level of G_0 .

Recall the notation introduced by the statement of Lemma 4.2. For given $r \geq 1$, for each $i = 1, \dots, k_0$, the system is given by the equations of real unknowns $(a_j, b_j, c_j, d_j)_{j=1}^{k-k_0+1}$:

$$\left\{ \sum_{j=1}^{k-k_0+1} \sum_{\alpha} \frac{P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_i^0)}{H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_i^0) Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v_i^0)} \frac{d_j^2 a_j^{\alpha_1} b_j^{\alpha_2} c_j^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!} = 0 \mid \beta \in \mathcal{F}_r \cap \{\beta_1 + 2\beta_2 + 2\beta_3 \leq r\} \right\} \quad (21)$$

where the range of $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{N}^3$ in the above sum satisfies $\alpha_1 + 2\alpha_2 + 2\alpha_3 = \beta_1 + 2\beta_2 + 2\beta_3$.

Note that the above system of polynomial equations is the general version of the systems of polynomial equations described in the previous section. There are $2r - 1$ equations in the above system of $4(k - k_0 + 1)$ unknowns. A solution of (21) is considered *non-trivial* if all of d_j are non-zeros while at least one among $a_1, \dots, a_i, b_1, \dots, b_i, c_1, \dots, c_i$ is non-zero. We say that system (21) is unsolvable if it does not have any non-trivial (or admissible) solution. The main result of this section is the following.

Theorem 4.1. *For each $i = 1, \dots, k_0$, let $\rho(v_i^0, m_i^0, k - k_0)$ be the minimum r for which system of polynomial equations (21) is unsolvable. Define*

$$R(G_0, k) = \max_{1 \leq i \leq k_0} \rho(v_i^0, m_i^0, k - k_0). \quad (22)$$

If $G_0 \in \mathcal{S}_0$, then $\ell(G_0|\mathcal{O}_{k,c_0}) \leq R(G_0, k) - 1$.

Remark: We make the following comments regarding the results of Theorem 4.1.

- (i) If $k - k_0 = 1$, we can obtain $R(G_0, k) = 4$ from the examples given in Section 4.1 (although in the examples we only worked out the case that $k_0 = 1$, for general $k_0 \geq 1$ the techniques are the same). Since G_0 is in fact 3-singular, the bound is tight.
- (ii) In order to determine $R(G_0, k)$, we need to find the value of $\rho(v_i^0, m_i^0, k - k_0)$ for all $1 \leq i \leq k_0$. One may ask whether the value of $\rho(v_i^0, m_i^0, k - k_0)$ depends on the specific values of v_i^0, m_i^0 . The structure of $\rho(v_i^0, m_i^0, k - k_0)$ will be looked at in more detail in the next subsection.

Proof. The strategy is clear: First, we shall obtain a valid r -minimal form for G_0 , cf. Eq. (8). This requires a method for obtaining linearly independent basis functions $H_l(x)$ out of the partial derivatives of kernel density f . Second, we obtain the polynomial limits of collection of coefficients of the r -minimal form. Third, we obtain bounds on r according to which this system of limiting polynomials does not admit non-trivial real solutions. This provides upper bounds on the singularity level of G_0 .

Step 1: Construction of r -minimal form It follows from Lemma 4.2 and Lemma 4.3 that a r -th minimal form for G_0 can be obtained as

$$\frac{p_G(x) - p_{G_0}(x)}{W_r^r(G, G_0)} \asymp \frac{A_1(x) + B_1(x)}{W_r^r(G, G_0)},$$

where $A_1(x)$ and $B_1(x)$ are given as follows

$$\begin{aligned} A_1(x) &= \sum_{i=1}^{k_0} \sum_{\beta \in \mathcal{F}_r} \left(\sum_{j=1}^{s_i} \sum_{|\alpha| \leq r} \frac{P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_i^0)}{H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_i^0) Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v_i^0)} \frac{p_{ij}(\Delta\theta_{ij})^{\alpha_1} (\Delta v_{ij})^{\alpha_2} (\Delta m_{ij})^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!} \right) \times \\ &\quad \frac{\partial^{|\beta|} f}{\partial \theta^{\beta_1} \partial v^{\beta_2} \partial m^{\beta_3}}(x | \theta_i^0, \sigma_i^0, m_i^0), \\ B_1(x) &= \sum_{i=1}^{k_0} \Delta p_i \cdot f(x | \theta_i^0, \sigma_i^0, m_i^0). \end{aligned}$$

Suppose that there exists a sequence of G tending to G_0 under W_r such that all the coefficients of $A_1(x)/W_r^r(G, G_0)$ and $B_1(x)/W_r^r(G, G_0)$ vanish. Then for all $1 \leq i \leq k_0$, we obtain that $\Delta p_i / W_r^r(G, G_0) \rightarrow 0$ and

$$E_{\beta_1, \beta_2, \beta_3}(\theta_i^0, v_i^0, m_i^0) := \frac{\sum_{j=1}^{s_i} \sum_{|\alpha| \leq r} \frac{P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_i^0)}{H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_i^0) Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v_i^0)} \frac{p_{ij}(\Delta\theta_{ij})^{\alpha_1} (\Delta v_{ij})^{\alpha_2} (\Delta m_{ij})^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!}}{W_r^r(G, G_0)} \rightarrow 0,$$

as $\beta \in \mathcal{F}_r$. According to Lemma 3.1, $W_r^r(G, G_0) \asymp D_r(G_0, G)$. So, $\sum_{i=1}^{k_0} |\Delta p_i| / D_r(G_0, G) \rightarrow 0$. It follows that

$$\sum_{i=1}^{k_0} \sum_{j=1}^{s_i} p_{ij} (|\Delta\theta_{ij}|^r + |\Delta v_{ij}|^r + |\Delta m_{ij}|^r) / D_r(G_0, G) \rightarrow 1.$$

This means there exists some index $i^* \in \{1, \dots, k_0\}$ such that

$$\sum_{j=1}^{s_{i^*}} p_{i^*j} (|\Delta\theta_{i^*j}|^r + |\Delta v_{i^*j}|^r + |\Delta m_{i^*j}|^r) / D_r(G_0, G) \not\rightarrow 0.$$

By multiplying the inverse of the above term with $E_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$ as $\beta \in \mathcal{F}_r$ and using the fact that $W_r^r(G, G_0) \asymp D_r(G_0, G)$, we obtain

$$\begin{aligned} F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0) &:= \\ &\frac{\sum_{j=1}^{s_1} \sum_{|\alpha| \leq r} \frac{P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_{i^*}^0)}{H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_{i^*}^0) Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v_{i^*}^0)} \frac{p_{i^*j}(\Delta\theta_{i^*j})^{\alpha_1} (\Delta v_{i^*j})^{\alpha_2} (\Delta m_{i^*j})^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!}}{\sum_{j=1}^{s_1} p_{i^*j} (|\Delta\theta_{i^*j}|^r + |\Delta v_{i^*j}|^r + |\Delta m_{i^*j}|^r)} \rightarrow 0, \end{aligned}$$

Step 2: Greedy extraction of polynomial limits We proceed to extract polynomial limits of all $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$. This technique has been demonstrated in Section 4.1 for specific cases. Note that the numerators of the $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$ are inhomogeneous polynomials in general. Let

$$\overline{M}_g = \max \left\{ |\Delta \theta_{i^*1}|, \dots, |\Delta \theta_{i^*s_{i^*}}|, |\Delta v_{i^*1}|^{1/2}, \dots, |\Delta v_{i^*s_{i^*}}|^{1/2}, |\Delta m_{i^*1}|^{1/2}, \dots, |\Delta m_{i^*s_{i^*}}|^{1/2} \right\}.$$

Denote the limits for the relevant subsequences, which exist due to the boundedness: $\Delta \theta_{i^*j}/\overline{M}_g \rightarrow a_j$, $\Delta v_{i^*j}/\overline{M}_g^2 \rightarrow b_j$, and $\Delta m_{i^*j}/\overline{M}_g^2 \rightarrow c_j$, and $p_{i^*j} \rightarrow d_j^2$ for each $j = 1, \dots, s_{i^*}$. Here, at least one element of $(a_j, b_j, c_j)_{j=1}^{s_{i^*}}$ equals to -1 or 1. For any index vector $\beta = (\beta_1, \beta_2, \beta_3)$ such that $\beta \in \mathcal{F}_r$, the lowest order of \overline{M}_g in the numerator of $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$ is $\overline{M}_g^{\beta_1 + 2\beta_2 + 2\beta_3}$. Since $\sum_{j=1}^{s_1} p_{i^*j}(|\Delta \theta_{i^*j}|^r + |\Delta v_{i^*j}|^r + |\Delta m_{i^*j}|^r) \asymp \overline{M}_g^r$, it is clear that $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$ vanishes as long as $\beta_1 + 2\beta_2 + 2\beta_3 \geq r + 1$. Thus, we only need to concern with $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$ when $\beta \in \mathcal{F}_r$ and $\beta_1 + 2\beta_2 + 2\beta_3 \leq r$.

For any $\beta = (\beta_1, \beta_2, \beta_3)$ such that $\beta \in \mathcal{F}_r$ and $\beta_1 + 2\beta_2 + 2\beta_3 \leq r$, by dividing the numerator and denominator of $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$ by $\overline{M}_g^{\beta_1 + 2\beta_2 + 2\beta_3}$ (i.e the lowest order of \overline{M}_g in the numerator of $F_{\beta_1, \beta_2, \beta_3}(\theta_{i^*}^0, v_{i^*}^0, m_{i^*}^0)$), we obtain the following system of equations

$$\sum_{j=1}^{s_{i^*}} \sum_{\alpha} \frac{P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_{i^*}^0)}{H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m_{i^*}^0) Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v_{i^*}^0)} \frac{d_j^2 a_j^{\alpha_1} b_j^{\alpha_2} c_j^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!} = 0, \quad (23)$$

where the range of $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ in the above sum satisfies $\alpha_1 + 2\alpha_2 + 2\alpha_3 = \beta_1 + 2\beta_2 + 2\beta_3$. The above system of polynomial equations is the general version of the systems of polynomial equations (19) and (20) that we considered in Section 4.1. Now, one of the elements of a_j s, b_j s, c_j s is non-zero.

Since $G \in \mathcal{O}_{k, c_0}$ and $\sum_{j=1}^{s_{i^*}} p_{i^*j} \rightarrow p_{i^*}^0$, we have the constraints $d_j^2 > 0$ and $\sum_{j=1}^{s_{i^*}} d_j^2 = p_{i^*}^0$. However,

we can remove the constraint on the summation of d_j^2 by putting $d_j^2 = p_{i^*}^0 (d'_j)^2 / \sum_{j=1}^{s_{i^*}} (d'_j)^2$ where

we the only constraint on d'_j s is $d'_j \neq 0$ for all $1 \leq j \leq s_{i^*}$. As a consequence, when we talk about system of polynomial equations (23), we can consider only the constraint $d_j^2 \neq 0$ for any $1 \leq j \leq s_{i^*}$. By Definition 3.2, G_0 is not r -singular relative to \mathcal{O}_{k, c_0} as long as the system (23) does not admit any non-trivial solution for the unknowns $(a_j, b_j, c_j, d_j)_{j=1}^{s_{i^*}}$.

Step 3: Deriving an upper bound There are two distinct features of system of polynomial equations (23). First, i^* varies in $\{1, 2, \dots, k_0\}$ as $G \in \mathcal{O}_{k, c_0}$ tends to G_0 . Second, the value of s_{i^*} of the subsequence of G is subject to the constraint that $s_{i^*} \leq k - k_0 + 1$. (This constraint arises due to number of distinct atoms of G , $\sum_{j=1}^{k_0} s_j \leq k' \leq k$ and all $s_j \geq 1$ for all $1 \leq j \leq k_0$). It follows from these two observations that the system (23) admits a non-trivial solution only if the system (21) also admits a non-trivial solution. This cannot be the case if $r \geq R(G_0, k)$, by the definition given in Eq. (22). This concludes our proof. \square

4.3 Properties of the system of limiting polynomial equations

The goal of this subsection is the present additional results on the structure of function $\rho(v, m, k - k_0)$, which is a fundamental quantity in Theorem 4.1 (Here, v_i^0, m_i^0 are replaced by v, m). It is difficult

to obtain explicit values for $\rho(v, m, k - k_0)$ in general. Nonetheless, we can obtain a nontrivial upper bound for ρ . Now, let $\Xi_1 := \{(v, m) \in \Theta_2 \times \Theta_3 : m \neq 0\}$. Recall that $\rho(v, m, l)$, where $l = k - k_0 \geq 1$, is the minimum value according to which system (21) does not admit non-trivial real-solution.

Proposition 4.1. *Let $\bar{r}(l)$ the minimal value of $s > 0$ such that the following system of polynomial equations*

$$\sum_{j=1}^l \sum_{\substack{n_1 + 2n_2 = \alpha \\ n_1, n_2 \geq 0}} \frac{x_j^2 y_j^{n_1} z_j^{n_2}}{n_1! n_2!} = 0 \text{ for each } \alpha = 1, \dots, s \quad (24)$$

does not have any solution for $(x_1, \dots, x_l, y_1, \dots, y_l, z_1, \dots, z_l)$ such that x_1, \dots, x_l are non-zeros, and at least one of y_1, \dots, y_l is non-zero. For all $l = 1, 2, \dots$, there holds

$$\sup_{(v, m) \in \Xi_1} \rho(v, m, l) \leq \bar{r}(l).$$

Remarks (i) The proof of this proposition is given in Appendix B, which proceeds by verifying that system (24) forms a subset of equations that define system (21). Combining with the statement of Theorem 4.1, we obtain

$$\ell(G_0 | \mathcal{O}_{k, c_0}) \leq \bar{r}(l) - 1.$$

(ii) A remarkable fact is that $\bar{r}(l)$ is nothing but the singularity level of G_0 relative to \mathcal{O}_{k, c_0} in the context of location-scale Gaussian mixture. This statement can be proved directly using the same method of proof described in the previous section for the skewnormal mixtures. The proof for the Gaussian mixture is much simpler, because the r -minimal form for Gaussian mixtures can be obtained via the relatively simpler elimination steps given by Example 3.3. The fact that the coefficients involved in this elimination are constant with respect to the model parameters is the fundamental reason why the singularity level of G_0 for the Gaussian mixtures is uniform over the entire space of parameters. See also Theorem 1.1 of Ho and Nguyen [2016a].

(iii) Combining the above remark with the results established by Theorem 3.2 leads us to conclude this: it is statistically more efficient to estimate location-scale-shape parameters of skewnormal o-mixtures than to estimate location-scale parameters of Gaussian o-mixtures that carry the same number of extra mixing components.

Dependence of ρ on (v, m) To understand the role of parameter value (v, m) on singularity levels, we shall construct a partition of the parameter space for (v, m) based on the value of function ρ . For each $l, r \geq 1$, define an “inverse” function

$$\rho_l^{-1}(r) = \{(v, m) \in \Xi_1 : \rho(v, m, l) = r\}.$$

Additionally, take

$$\underline{\rho}(l) = \min \{r : \rho_l^{-1}(r) \neq \emptyset\}, \quad \bar{\rho}(l) = \max \{r : \rho_l^{-1}(r) \neq \emptyset\}.$$

It follows from Proposition 4.1 that $\bar{\rho}(l) \leq \bar{r}(l)$. In addition, $\rho_l^{-1}(r)$ are mutually disjoint for different values of r . So, for each fixed $l \geq 1$,

$$\Xi_1 = \bigcup_{r=\underline{\rho}(l)}^{\bar{\rho}(l)} \rho_l^{-1}(r).$$

Proposition 4.2. *For each $l \geq 1, r \geq 1$, $\rho_l^{-1}(r)$ is a semialgebraic set.*

Proof. For each $r \geq 1$, let \mathbb{A}_r be the collection of all $(v, m) \in \Xi_1$ such that the system of polynomial equations (21) contains admissible solutions. Furthermore, \mathbb{B}_r denotes the collection of all solutions $(v, m, \{a_i\}_{i=1}^l, \{b_i\}_{i=1}^l, \{c_i\}_{i=1}^l, \{d_i\}_{i=1}^l)$ of system of polynomial equations (21), i.e., we treat v, m as two additional unknowns of the system. Since $P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m)$, $H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m)$, and $Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v)$ are polynomial functions of m, v for all α, β , by definition \mathbb{B}_r is a semialgebraic set for all $r \geq 1$. By Tarski-Seidenberg theorem [Basu et al., 2006], since \mathbb{A}_r is the projection of \mathbb{B}_r from dimension $(4l+2)$ to dimension 2, \mathbb{A}_r is a semialgebraic set for all $r \geq 1$. It follows that \mathbb{A}_r^c is semialgebraic for all $r \geq 1$. Since $\rho_l^{-1}(r) = \mathbb{A}_r^c \cap \mathbb{A}_{r-1}$ for all $r \geq 1$, the conclusion of the proposition follows. \square

The following result gives us some exact values of $\underline{\rho}(l)$ and $\overline{\rho}(l)$ in specific cases.

Proposition 4.3. (a) *If $l = k - k_0 = 1$, then $\underline{\rho}(l) = \overline{\rho}(l) = 4$.*

(b) *If $l = k - k_0 = 2$, then $\underline{\rho}(l) = 5$ and $\overline{\rho}(l) = 6$. Thus, Ξ_1 is partitioned into two subsets, both of which are non-empty because $\{(1, -2), (1, 2)\} \subset \rho_l^{-1}(5)$, and $(1, \frac{1}{10}) \in \rho_l^{-1}(6)$.*

From the definition of $R(G_0, k)$, we can write

$$R(G_0, k) = \max \left\{ r \mid \text{there is } i = 1, \dots, k_0 \text{ such that } (v_i^0, m_i^0) \in \rho_{k-k_0}^{-1}(r) \right\}.$$

According to the Proposition 4.3, if $k - k_0 = 1$, we have $R(G_0, k) = 4$ (see also our earlier remark). If $k - k_0 = 2$, we may have either $R(G_0, k) = 5$ or 6, depending on the value of parameters (v, m) that provide the support for G_0 .

We end this subsection by noting that we have just provided specific examples in which $R(G_0, k) - 1$ may vary with the actual parameter values that define G_0 . Although this is an upper bound of the singularity level, we have *not* actually proved that the singularity level of G_0 may generally vary with its parameter values. We will be able to do so in the sequel, when we work with the e-mixture setting.

5 E-mixtures of skewnormal distributions

E-mixtures are the setting in which the number of mixing components is known $k = k_0$. In this section, we study the singularity structure of mixing measure G_0 relative to the ambient space \mathcal{E}_{k_0} , where k_0 is the number of supporting atoms for G_0 .

Recall from the previous section the definition of \mathcal{S}_0 , the subset $\mathcal{S}_0 \subset \mathcal{E}_{k_0}$ of measure $G_0 = G_0(\mathbf{p}^0, \boldsymbol{\eta}^0)$ such that $(\mathbf{p}^0, \boldsymbol{\eta}^0)$ satisfy $P_1(\boldsymbol{\eta}^0)P_2(\boldsymbol{\eta}^0) \neq 0$. P_1 and P_2 are polynomials given in the statement of Lemma 4.1. It is simple to verify that for any $G_0 \in \mathcal{S}_0$, as a consequence of this lemma, the Fisher information matrix $I(G_0)$ is non-singular. It follows that

Theorem 5.1. *If $G_0 \in \mathcal{S}_0$, then $\ell(G_0 | \mathcal{E}_{k_0}) = 0$.*

We turn our attention to the singularity structure of set $\mathcal{E}_{k_0} \setminus \mathcal{S}_0$. For any $G_0 \in \mathcal{E}_{k_0} \setminus \mathcal{S}_0$, the parameters of G_0 satisfy $P_1(\boldsymbol{\eta}^0)P_2(\boldsymbol{\eta}^0) = 0$. Accordingly, for each pair of $(i, j) = 1, \dots, k_0$ the two components indexed by i and j are said to be **homologous** if

$$(\theta_i^0 - \theta_j^0)^2 + [v_i^0(1 + (m_j^0)^2) - v_j^0(1 + (m_i^0)^2)]^2 = 0.$$

Moreover, for each $1 \leq i \leq k_0$, let I_i denote the set of all components homologous to (component) i . By definition, it is clear that if i and j are homologous, $I_i \equiv I_j$. Therefore, these homologous sets form

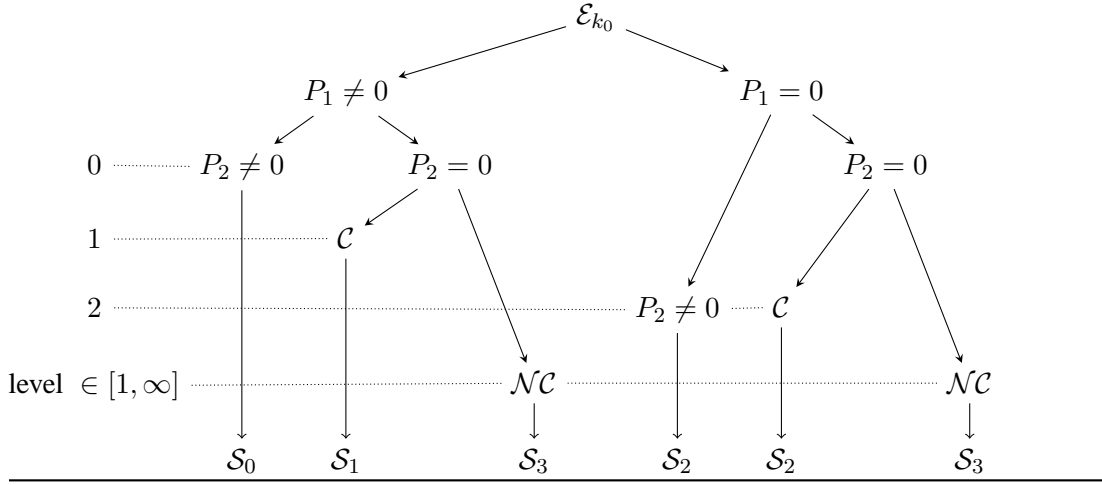


Figure 2: The singularity level of G_0 relative to \mathcal{E}_{k_0} is determined by partition based on zeros of polynomials P_1, P_2 into subsets $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$. Here, " \mathcal{NC} " stands for nonconformant.

equivalence classes. From here on, when we say a homologous set I , we implicitly mean that it is the representation of the equivalent classes.

Now, the homologous set consists of the indices of skewnormal components that share the same location and a rescaled version of the scale parameter. A non-empty homologous set I is said to be **conformant** if for any $i \neq j \in I$, $m_i^0 m_j^0 > 0$. A non-empty homologous set I is said to be **nonconformant** if we can find two indices $i, j \in I$ such that $m_i^0 m_j^0 < 0$. Additionally, G_0 is said to be **conformant** if all the homologous sets are conformant or **nonconformant** (NC) if at least one homologous set is nonconformant. Now, we define a partition of $\mathcal{E}_{k_0} \setminus \mathcal{S}_0$ as follows $\mathcal{E}_{k_0} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$, where

$$\begin{cases} \mathcal{S}_1 = \{G = G(\mathbf{p}, \boldsymbol{\eta}) \in \mathcal{E}_{k_0} \mid P_1(\boldsymbol{\eta}) \neq 0, P_2(\boldsymbol{\eta}) = 0, G \text{ is conformant}\} \\ \mathcal{S}_2 = \{G = G(\mathbf{p}, \boldsymbol{\eta}) \in \mathcal{E}_{k_0} \mid P_1(\boldsymbol{\eta}) = 0, \text{ if } P_2(\boldsymbol{\eta}) = 0 \text{ then } G \text{ is conformant}\} \\ \mathcal{S}_3 = \{G = G(\mathbf{p}, \boldsymbol{\eta}) \in \mathcal{E}_{k_0} \mid P_2(\boldsymbol{\eta}) = 0 \text{ and } G \text{ is nonconformant}\}. \end{cases}$$

Figure 2 summarizes singularity levels of elements residing in \mathcal{E}_{k_0} , except for \mathcal{S}_3 .

5.1 Singularity level of $G_0 \in \mathcal{S}_1 \cup \mathcal{S}_2$

The main results of this subsection are the following two theorems

Theorem 5.2. *If $G_0 \in \mathcal{S}_1$, then $\ell(G_0 | \mathcal{E}_{k_0}) = 1$.*

Theorem 5.3. *If $G_0 \in \mathcal{S}_2$, then $\ell(G_0 | \mathcal{E}_{k_0}) = 2$.*

The complete proofs of both theorems are given in in Appendix. In the following, we shall present the proof for a simple setting of $G_0 \in \mathcal{S}_1$, which illustrates the complete proofs, and also helps to explain why the partition of according to \mathcal{S}_1 , i.e., the notion of conformant, arises in the first place.

Proof. (of a simplified setting) The simplified setting is that all components of G_0 are homologous to one another. By definition all components of G_0 are non-Gaussian (because $P_1(\boldsymbol{\eta}^0) \neq 0$). Thus, we have $\theta_1^0 = \dots = \theta_{k_0}^0$ and $\frac{v_1^0}{1 + (m_1^0)^2} = \dots = \frac{v_{k_0}^0}{1 + (m_{k_0}^0)^2}$. Additionally, $m_i^0 \neq 0$ for all $1 \leq i \leq k_0$.

Since G_0 is conformant, m_i^0 share the same sign for all $1 \leq i \leq k_0$. Without loss of generality, we assume $m_i^0 > 0$. We need to show that G_0 is 1-singular, but not 2-singular.

G_0 is 1-singular Given constraints on the parameters of G_0 , it is simple to arrive at the following 1-minimal form (cf. Eq. (8)):

$$\begin{aligned} \frac{1}{W_1(G, G_0)} \left\{ \sum_{i=1}^{k_0} \left[\beta_{1i}^{(1)} + \beta_{2i}^{(1)}(x - \theta_1^0) + \beta_{3i}^{(1)}(x - \theta_1^0)^2 \right] f\left(\frac{x - \theta_1^0}{\sigma_i^0}\right) \Phi\left(\frac{m_i^0(x - \theta_1^0)}{\sigma_i^0}\right) \right. \\ \left. + \left[\gamma_1^{(1)} + \gamma_2^{(1)}(x - \theta_1^0) \right] \exp\left(-\frac{(m_1^0)^2 + 1}{2v_1^0}(x - \theta_1^0)^2\right) \right\} + o(1), \quad (25) \end{aligned}$$

where coefficients $\beta_{1i}^{(1)}, \beta_{2i}^{(1)}, \beta_{3i}^{(1)}, \gamma_1^{(1)}, \gamma_2^{(1)}$ are the polynomials of $\Delta\theta_j, \Delta v_j, \Delta m_j$, and Δp_j :

$$\begin{aligned} \beta_{1i}^{(1)} &= \frac{2\Delta p_i}{\sigma_i^0} - \frac{p_i \Delta v_i}{(\sigma_i^0)^3}, \quad \beta_{2i}^{(1)} = \frac{2p_i \Delta \theta_i}{(\sigma_i^0)^3}, \quad \beta_{3i}^{(1)} = \frac{p_i \Delta v_i}{(\sigma_i^0)^5}, \\ \gamma_1^{(1)} &= \sum_{j=1}^{k_0} -\frac{p_j m_j^0 \Delta \theta_j}{\pi(\sigma_j^0)^2}, \quad \gamma_2^{(1)} = \sum_{j=1}^{k_0} -\frac{p_j m_j^0 \Delta v_j}{2\pi(\sigma_j^0)^4} + \frac{p_j \Delta m_j}{\pi(\sigma_j^0)^2}. \end{aligned}$$

Note that, the conditions $m_i^0 \neq 0$ for all $1 \leq i \leq k_0$ allow us to have that $f\left(\frac{x - \theta_1^0}{\sigma_i^0}\right) \Phi\left(\frac{m_i^0(x - \theta_1^0)}{\sigma_i^0}\right)$ and $\exp\left(-\frac{(m_1^0)^2 + 1}{2v_1^0}(x - \theta_1^0)^2\right)$ are linearly independent. It is clear that if a sequence of G (represented by Eq. (4)) is chosen such that $\Delta\theta_i = \Delta v_i = \Delta p_i = 0$ for all $1 \leq i \leq k_0$, and $\sum_{i=1}^{k_0} p_i \Delta m_i / v_i^0 = 0$, then we obtain $\beta_{1i}^{(1)} / W_1(G, G_0) = \beta_{2i}^{(1)} / W_1(G, G_0) = \beta_{3i}^{(1)} / W_1(G, G_0) = \gamma_1^{(1)} / W_1(G, G_0) = \gamma_2^{(1)} / W_1(G, G_0) = 0$. Hence, G_0 is 1-singular relative to \mathcal{E}_{k_0} .

G_0 is not 2-singular Indeed, suppose that this is not true. Then from Definition 3.2, for any sequence of G that tends to G_0 under W_2 , all coefficients of the 2-minimal form must vanish. A 2-minimal form is given as follows:

$$\begin{aligned} \frac{1}{W_2^2(G, G_0)} \left[\sum_{i=1}^{k_0} \left(\sum_{j=1}^5 \beta_{ji}^{(2)}(x - \theta_1^0)^{j-1} \right) f\left(\frac{x - \theta_1^0}{\sigma_i^0}\right) \Phi\left(\frac{m_i^0(x - \theta_1^0)}{\sigma_i^0}\right) \right. \\ \left. + \left(\sum_{j=1}^4 \gamma_j^{(2)}(x - \theta_1^0)^{j-1} \right) \exp\left(-\frac{(m_1^0)^2 + 1}{2v_1^0}(x - \theta_1^0)^2\right) \right] + o(1), \quad (26) \end{aligned}$$

where $\beta_{ji}^{(2)}, \gamma_j^{(2)}$ are polynomials of $\Delta\theta_l, \Delta v_l, \Delta m_l$, and Δp_l for $l = 1, \dots, k_0$:

$$\begin{aligned}\beta_{1i}^{(2)} &= \frac{2\Delta p_i}{\sigma_i^0} - \frac{p_i \Delta v_i}{(\sigma_i^0)^3} - \frac{p_i (\Delta\theta_i)^2}{(\sigma_i^0)^3} + \frac{3p_i (\Delta v_i)^2}{4(\sigma_i^0)^5}, \quad \beta_{2i}^{(2)} = \frac{2p_i \Delta\theta_i}{(\sigma_i^0)^3} - \frac{6p_i \Delta\theta_i \Delta v_i}{(\sigma_i^0)^5}, \\ \beta_{3i}^{(2)} &= \frac{p_i \Delta v_i}{(\sigma_i^0)^5} + \frac{p_i (\Delta\theta_i)^2}{(\sigma_i^0)^5} - \frac{3p_i (\Delta v_i)^2}{2(\sigma_i^0)^7}, \quad \beta_{4i}^{(2)} = \frac{2p_i \Delta\theta_i \Delta v_i}{(\sigma_i^0)^7}, \quad \beta_{5i}^{(2)} = \frac{p_i (\Delta v_i)^2}{4(\sigma_i^0)^9}, \\ \gamma_1^{(2)} &= \sum_{j=1}^{k_0} -\frac{p_j m_j^0 \Delta\theta_j}{\pi(\sigma_j^0)^2} + \frac{2p_j m_j^0 (\Delta\theta_j)(\Delta v_j)}{\pi(\sigma_j^0)^4} - \frac{2p_j \Delta\theta_j \Delta m_j}{\pi(\sigma_j^0)^2}, \\ \gamma_2^{(2)} &= \sum_{j=1}^{k_0} -\frac{p_j m_j^0 \Delta v_j}{2\pi(\sigma_j^0)^4} - \frac{p_j ((m_j^0)^3 + 2m_j^0)(\Delta\theta_j)^2}{2\pi(\sigma_j^0)^4} + \frac{p_j \Delta m_j}{\pi(\sigma_j^0)^2} + \frac{5p_j m_j^0 (\Delta v_j)^2}{8\pi(\sigma_j^0)^6} - \frac{p_j \Delta v_j \Delta m_j}{\pi(\sigma_j^0)^4}, \\ \gamma_3^{(2)} &= \sum_{j=1}^{k_0} \frac{p_j (2(m_j^0)^2 + 2)\Delta\theta_j \Delta m_j}{\pi(\sigma_j^0)^4} - \frac{p_j ((m_j^0)^3 + 2m_j^0)\Delta\theta_j \Delta v_j}{2\pi(\sigma_j^0)^6}, \\ \gamma_4^{(2)} &= \sum_{j=1}^{k_0} -\frac{p_j ((m_j^0)^3 + 2m_j^0)(\Delta v_j)^2}{8\pi(\sigma_j^0)^8} - \frac{p_j m_j^0 (\Delta m_j)^2}{2\pi(\sigma_j^0)^4} + \frac{p_j ((m_j^0)^2 + 1)\Delta v_j \Delta m_j}{\pi(\sigma_j^0)^6}.\end{aligned}$$

Now, $\beta_{ji}^{(2)}/W_2^2(G, G_0) \rightarrow 0$ leads to $\Delta p_i/W_2^2(G, G_0), \Delta\theta_i/W_2^2(G, G_0), \Delta v_i/W_2^2(G, G_0) \rightarrow 0$ for all $1 \leq i \leq k_0$ (The rigorous argument for that result is in Step 1.1 of the full proof of this theorem in Appendix B). Combining with Lemma 3.1, we obtain

$$\sum_{i=1}^{k_0} p_i |\Delta m_i|^2 / W_2^2(G, G_0) \not\rightarrow 0. \quad (27)$$

Additionally, the vanishing of coefficients $\gamma_j^{(2)}/W_2^2(G, G_0)$ for $1 \leq j \leq 4$ entails

$$\begin{aligned}\left(\sum_{i=1}^{k_0} p_i \Delta m_i / v_i^0 \right) / W_2^2(G, G_0) &\rightarrow 0, \\ \left(\sum_{i=1}^{k_0} p_i m_i^0 (\Delta m_i)^2 / (v_i^0)^2 \right) / W_2^2(G, G_0) &\rightarrow 0.\end{aligned} \quad (28)$$

Combining (27) and (28), it follows that

$$\left(\sum_{i=1}^{k_0} p_i m_i^0 (\Delta m_i)^2 / (v_i^0)^2 \right) / \sum_{i=1}^{k_0} p_i |\Delta m_i|^2 \rightarrow 0,$$

which is a contradiction due to $m_i^0 > 0$ for all $1 \leq i \leq k_0$. Hence, G_0 is not 2-singular relative to \mathcal{E}_{k_0} . We conclude that $\ell(G_0|\mathcal{E}_{k_0}) = 1$. \square

5.2 Singularity levels of $G_0 \in \mathcal{S}_3$: a summary

The singularity structure of \mathcal{S}_3 is much more complex than those of previous settings of G_0 . \mathcal{S}_3 does not admit an uniform level of singularity for all its elements — it needs to be partitioned into many subsets via intersections with additional semialgebraic sets of the parameter space. In addition, we can

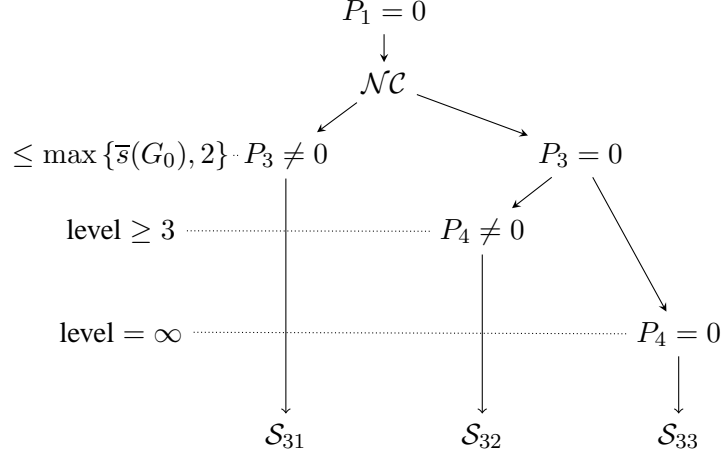


Figure 3: The level of singularity structure of $G_0 \in \mathcal{S}_3$ when $P_1(\eta^0) = 0$. Here, "NC" stands for nonconformant. The term $\bar{s}(G_0)$ is defined in (37).

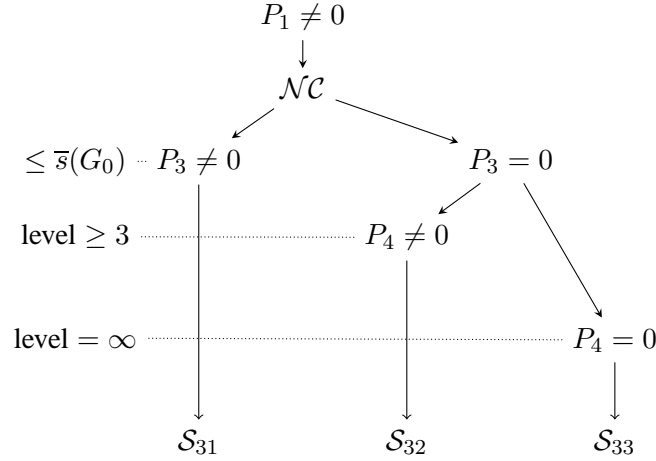


Figure 4: The level of singularity structure of $G_0 \in \mathcal{S}_3$ when $P_1(\eta^0) \neq 0$. Here, "NC" stands for nonconformant. The term $\bar{s}(G_0)$ is defined in (37).

establish the existence of subsets that correspond to the infinite level of singularity. In most cases when the singularity level is finite, we may only be able to provide some bounds rather than an exact values. As in o-mixtures, the unifying theme of such bounds is their connection to the solvability of a system of real polynomial equations.

If $G_0 = G_0(\mathbf{p}^0, \boldsymbol{\eta}^0) \in \mathcal{S}_3$, then its corresponding parameters satisfy $P_2(\boldsymbol{\eta}^0) = 0$, i.e., there is at least one homologous set of G_0 . Moreover, at least one such homologous set is nonconformant. For any $G_0 \in \mathcal{S}_3$, let I_1, \dots, I_t be all nonconformant homologous sets of G_0 . The singularity structures of \mathcal{S}_3 arise from the zeros of the following polynomials:

- Type C(1): $P_3(\mathbf{p}^0, \boldsymbol{\eta}^0) := \prod_{i=1}^t \left(\prod_{S \subseteq I_i, |S| \geq 2} \left(\sum_{j \in S} p_j^0 \prod_{l \neq j} m_l^0 \right) \right)$.
- Type C(2): $P_4(\mathbf{p}^0, \boldsymbol{\eta}^0) := \prod_{1 \leq i \neq j \leq k_0} \left[u_{ij}^2 + (m_i^0 \sigma_j^0 + m_j^0 \sigma_i^0)^2 + (p_i^0 \sigma_j^0 - p_j^0 \sigma_i^0)^2 \right]$, where $u_{ij}^2 = (\theta_i^0 - \theta_j^0)^2 + \left(v_i^0(1 + (m_j^0)^2) - v_j^0(1 + (m_i^0)^2) \right)^2$.

Type C singularities, including both C(1) and C(2), are distinguished from Type A and Type B singularities by the fact that the Type C polynomials are defined by not only component parameters $\boldsymbol{\eta}^0$, but also mixing probability parameters \mathbf{p}^0 . Note that C(1) singularity implies that there is some homologous set I_i of G_0 such that $\prod_{S \subseteq I_i, |S| \geq 2} \left(\sum_{j \in S} p_j^0 \prod_{l \neq j} m_l^0 \right) = 0$. A homologous set of G_0 having the above property is said to contain type C(1) singularity locally. Similarly, type C(2) singularity implies that there is some pair $1 \leq i \neq j \leq k_0$ such that $u_{ij}^2 + (m_i^0 \sigma_j^0 + m_j^0 \sigma_i^0)^2 + (p_i^0 \sigma_j^0 - p_j^0 \sigma_i^0)^2 = 0$. A homologous set of G_0 having this pair is said to contain type C(2) singularity locally. It can be easily checked that a homologous set containing type C(2) singularity must also contain type C(1) singularity, since $P_4(\mathbf{p}^0, \boldsymbol{\eta}^0) = 0$ entails $P_3(\mathbf{p}^0, \boldsymbol{\eta}^0) = 0$. Now, we define the following partition of \mathcal{S}_3 according to the definition of type C(1) and C(2) singularity: $\mathcal{S}_3 = \mathcal{S}_{31} \cup \mathcal{S}_{32} \cup \mathcal{S}_{33}$, where

$$\begin{cases} \mathcal{S}_{31} = \{G = G(\mathbf{p}, \boldsymbol{\eta}) \in \mathcal{S}_3 \mid P_3(\mathbf{p}, \boldsymbol{\eta}) \neq 0\} \\ \mathcal{S}_{32} = \{G = G(\mathbf{p}, \boldsymbol{\eta}) \in \mathcal{S}_3 \mid P_3(\mathbf{p}, \boldsymbol{\eta}) = 0, P_4(\mathbf{p}, \boldsymbol{\eta}) \neq 0\} \\ \mathcal{S}_{33} = \{G = G(\mathbf{p}, \boldsymbol{\eta}) \in \mathcal{S}_3 \mid P_3(\mathbf{p}, \boldsymbol{\eta}) = 0, P_4(\mathbf{p}, \boldsymbol{\eta}) = 0\}. \end{cases}$$

Due to the highly technical nature of our analysis of the singularity structure of \mathcal{S}_3 , we defer such details to Section 7.1 in Appendix A. Here, we only provide a summary of such results. Figure 3 and 4 provide additional illustrations. Specifically, when $G_0 \in \mathcal{S}_{31}$, it is shown that $\ell(G_0 | \mathcal{E}_{k_0}) \leq \max\{2, \bar{\mathfrak{s}}(G_0)\}$, where $\bar{\mathfrak{s}}(G_0)$ is defined by a system of polynomial equations that we obtain via a method of greedy extraction of polynomial limits, see Section 7.1.1. In some specific cases, the precise singularity level of $G_0 \in \mathcal{S}_{31}$ will be given. If $G_0 \in \mathcal{S}_{32}$, we need a more sophisticated method of extraction for polynomial limits; our technique is illustrated on a specific example of G_0 in Section 7.1.2. Finally, if $G_0 \in \mathcal{S}_{33}$, it is shown that $\ell(G_0 | \mathcal{E}_{k_0}) = \infty$ in Section 7.1.3.

6 Discussion and concluding remarks

Understanding the behavior of parameter estimates of mixture models is useful because the mixing parameters represent explicitly the heterogeneity of the underlying data population that mixture models are most suitable for. In this paper, a general framework for the identification of singularity structure arising from finite mixture models is proposed. It is shown that the singularity levels of the model's

parameter space directly determine minimax lower bounds and maximum likelihood estimation convergence rates, under conditions on the compactness of the parameter space.

The systematic identification of singularity levels and the implications on parameter estimation is a crucial step toward the development of more efficient model-based inference procedures. It is our view that such procedures must account for the presence of singular points residing in the parameter space of the model. As a matter of fact, there are quite a few examples of such efforts applied to specific statistical models, even if the picture of the singularity structure associating with those models might not have been discussed explicitly. This raises a question of whether or not it is possible to extend and generalize such techniques in order to address the presence of singularities in a direct fashion. We give several examples:

- (1) For overfitted mixture models, methods based on likelihood-based penalization techniques were shown to be quite effective (e.g., [Toussile and Gassiat, 2009, Chen, 2016]). Our work shows that parameter values residing in the vicinity of regions of high singularity levels should be hard to estimate efficiently. Can a penalization technique be generalized to regularize the estimates toward subsets containing singularity points of smaller levels?
- (2) Suitable choices of Bayesian prior have been proposed to induce favorable posterior contraction behavior for overfitted finite mixtures [Rousseau and Mengersen, 2011]. Can we develop an appropriate prior for the mixture model parameters, given our knowledge of singular points residing in the parameter space?
- (3) Reparametrization is an effective technique that can be employed to combat singularities present in the class of skewed distributions [Hallin and Ley, 2014]. It would be interesting to study if such reparameterization technique can be systematically developed for the mixture models as well.

We also expect that the theory of singularity structures carries important consequences on the computational complexity of parameter estimation procedures, including both optimization and sampling based methods. The non-uniform nature of the singularity levels reveals a complex structure of the likelihood function: regions in parameter space that carry low singularity levels may observe a relatively high curvature of the likelihood surface, while high singularity levels imply a “flatter” likelihood surface along a certain subspace of the parameters. Such a subspace is manifested by our construction of sequences of mixing measures that attest to the condition of r -singularity. It is of interest to exploit the explicit knowledge of singularity levels obtained for a given mixture model class, so as to improve upon the computational efficiency of the optimization and sampling procedures that operate on the model’s parameter space.

References

- E. S. Allman, C. Matias, and J. A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics*, 37:3099–3132, 2009.
- R. B. Arellano-Valle, L. M. Castro, M. C. Genton, and H. W. Gómez. Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis*, 3: 513–540, 2008.
- R. B. Arellano-Valle, M. C. Genton, and R. H. Loschi. Shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis*, 100:91–101, 2009.

- A. Azzalini. Further results on a class of distributions which includes the normal ones. *Statistica (Bologna)*, 46:199–208, 1986.
- A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B(Methodological)*, 61:579–602, 1999.
- A. Azzalini and A. D. Valle. The multivariate skew-normal distribution. *Biometrika*, 83:715–726, 1996.
- S. Basu, R. Pollack, and M. Roy. *Algorithms in real algebraic geometry*. Springer-Verlag Berlin Heidelberg, 2006.
- B. Buchberger. *An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal*. PhD thesis, Johannes Kepler University of Linz, 1965.
- P. Bühlmann and S. van de Geer. *Statistics for high dimensional data*. Springer, 2011.
- A. Canale and B. Scarpa. Bayesian nonparametric location-scale-shape mixtures. *TEST*, pages 1–18, 2015.
- H. Chen and J. Chen. Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, 13:351–365, 2003.
- J. Chen. Consistency of the mle under mixture models. *arXiv preprint arXiv:1607.01251*, 2016.
- J. H. Chen. Optimal rate of convergence for finite mixture models. *Annals of Statistics*, 23(1):221–233, 1995.
- M. Chiogna. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods and Applications*, 14:331–341, 2005.
- D. Cox, J. Little, and D. O’Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2007.
- A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer, 2008.
- M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on Algebraic Statistics*. Birkhauser, 2009.
- S. Ghosal and A. Roy. Predicting false discovery proportion under dependence. *Journal of the American Statistical Association*, 106:1208–1217, 2011.
- S. Ghosal and A. van der Vaart. Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities. *Annals of Statistics*, 29:1233–1263, 2001.
- M. Hallin and C. Ley. Skew-symmetric distributions and fisher information - a tale of two densities. *Bernoulli*, 18:747–763, 2012.
- M. Hallin and C. Ley. Skew-symmetric distributions and fisher information: the double sin of skew-normal. *Bernoulli*, 20:1432–1453, 2014.
- T. Hastie, R. Tibshirani, and M. J. Wainwright. *Statistical Learning with Sparsity: The Lasso and generalizations*. CRC Press, 2015.
- N. Ho and X. Nguyen. Convergence rates of parameter estimation for some weakly identifiable finite mixtures. *Annals of Statistics*, 2016a.

- N. Ho and X. Nguyen. On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electronic Journal of Statistics*, 10:271–307, 2016b.
- N. Ho and X. Nguyen. Singularity structures and impacts on parameter estimation in finite mixtures of distributions. Technical Report 540, Department of Statistics, University of Michigan, 2016c.
- H. Kasahara and K. Shimotsu. Testing the number of components in normal mixture regression models. *Journal of the American Statistical Association*, 2014.
- L. F. Lee and A. Chesher. Specification testing when score test statistics are identically zero. *Journal of Econometrics*, 31:33–61, 1986.
- S. X. Lee and G. J. McLachlan. On mixtures of skew normal and skew t -distributions. *Advances in Data Analysis and Classification*, 7:241–266, 2013.
- E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, 1998.
- C. Ley and D. Paindaveine. On the singularity of multivariate skew-symmetric models. *Journal of Multivariate Analysis*, 101:1434–1444, 2010.
- T. I. Lin. Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, 100:257–265, 2009.
- T. I. Lin, J. C. Lee, and S. Y. Yen. Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, 17:909–927, 2007.
- B. Lindsay. *Mixture models: Theory, geometry and applications*. In NSF-CBMS Regional Conference Series in Probability and Statistics. IMS, Hayward, CA., 1995.
- X. Nguyen. Convergence of latent mixing measures in finite and infinite mixture models. *Annals of Statistics*, 4(1):370–400, 2013.
- M. O. Prates, C. R. B. Cabral, and V. H. Lachos. mixsmsn: fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, 54, 2013.
- A. Rotnitzky, D. R. Cox, M. Bottai, and J. Robins. Likelihood-based inference with singular information matrix. *Bernoulli*, 6:243–284, 2000.
- J. Rousseau and K. Mengersen. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B*, 73(5):689–710, 2011.
- S. W. Schnatter and S. Pyne. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, 11:317–336, 2009.
- B. Stumfel. *Solving systems of polynomial equations*. Providence, RI: American Mathematical Soc, 2002.
- W. Toussile and E. Gassiat. Variable selection in model-based clustering using multilocus genotype data. *Advances in Data Analysis and Classification*, 3:109–134, 2009.
- S. van de Geer. *Empirical Processes in M-estimation*. Cambridge University Press, 2000.
- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

- S. Xiao and G. Zeng. Determination of the limits for multivariate rational functions. *Science China Mathematics*, 57:397–416, 2014.
- C. B. Zeller, C. R. B. Cabral, and V. H. Lachos. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *TEST*, pages 1–22, 2015.

7 Appendix A

This Appendix contains additional results on the singularity structure of e-mixtures of skewnormal distributions.

7.1 Singularity structure of \mathcal{S}_3 : detailed analysis

To develop intuition and obtain bounds for singularity level for $G_0 \in \mathcal{S}_3$, we start by considering a simple case similar to the exposition of subsection 5.1. That is, G_0 has only one homologous set of size k_0 . $G_0 \in \mathcal{S}_3$ means that m_i^0 do not share the same signs for all $i = 1, \dots, k_0$. To investigate the singularity level for G_0 , we first obtain an r -minimal form, for $r \geq 2$, of $(p_G(x) - p_{G_0}(x))/W_r^r(G, G_0)$ by

$$\begin{aligned} \frac{1}{W_r^r(G, G_0)} & \left[\sum_{i=1}^{k_0} \left(\sum_{j=1}^{2r+1} \beta_{ji}^{(r)} (x - \theta_1^0)^{j-1} \right) f\left(\frac{x - \theta_1^0}{\sigma_i^0}\right) \Phi\left(\frac{m_i^0(x - \theta_1^0)}{\sigma_i^0}\right) \right. \\ & \left. + \left(\sum_{j=1}^{2r} \gamma_j^{(r)} (x - \theta_1^0)^{j-1} \right) \exp\left(-\frac{(m_1^0)^2 + 1}{2v_1^0} (x - \theta_1^0)^2\right) \right] + o(1), \end{aligned} \quad (29)$$

where $\beta_{ji}^{(r)}, \gamma_j^{(r)}$ are polynomials of $\Delta\theta_l, \Delta v_l, \Delta m_l$, and Δp_l as $1 \leq i, l \leq k_0$ and $1 \leq j \leq 2r + 1$. For concrete formulas of $\beta_{ji}^{(r)}, \gamma_j^{(r)}$, we note that for any $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ such that $|\alpha| \leq r$, there holds

$$\begin{aligned} \frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} \partial v^{\alpha_2} \partial m^{\alpha_3}} &= \left(\sum_{i=1}^{2r} \frac{U_i^{\alpha_1, \alpha_2, \alpha_3}(m)}{V_i^{\alpha_1, \alpha_2, \alpha_3}(v)} (x - \theta)^{i-1} \right) f\left(\frac{x - \theta}{\sigma}\right) f\left(\frac{m(x - \theta)}{\sigma}\right) + \\ & \frac{1}{\sigma} \left(\sum_{i=1}^{2r+1} \frac{L_i^{\alpha_1, \alpha_2, \alpha_3}}{N_i^{\alpha_1, \alpha_2, \alpha_3}(v)} (x - \theta)^{i-1} \right) f\left(\frac{x - \theta}{\sigma}\right) \Phi\left(\frac{m(x - \theta)}{\sigma}\right). \end{aligned}$$

In the above display $U_i^{\alpha_1, \alpha_2, \alpha_3}(m), V_i^{\alpha_1, \alpha_2, \alpha_3}(v), N_i^{\alpha_1, \alpha_2, \alpha_3}(v)$ are polynomials in terms of m, v and $L_i^{\alpha_1, \alpha_2, \alpha_3}$ are some constant numbers. As $\alpha_3 \geq 1$, we can further check that $L_i^{\alpha_1, \alpha_2, \alpha_3} = 0$ for all $1 \leq i \leq 2r$ and α_1, α_2 such that $|\alpha| \leq r$. It follows that

$$\begin{aligned} \beta_{ji}^{(r)} &= \frac{2\Delta p_i}{\sigma_j^0} 1_{\{j=1\}} + \frac{1}{\sigma_i^0} \sum_{|\alpha| \leq r} \frac{L_j^{\alpha_1, \alpha_2, \alpha_3}}{N_j^{\alpha_1, \alpha_2, \alpha_3}(v_i^0)} \frac{p_i(\Delta\theta_i)^{\alpha_1} (\Delta v_i)^{\alpha_2} (\Delta m_i)^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!}, \\ \gamma_j^{(r)} &= \sum_{i=1}^{k_0} \sum_{|\alpha| \leq r} \frac{U_j^{\alpha_1, \alpha_2, \alpha_3}(m_i^0)}{V_j^{\alpha_1, \alpha_2, \alpha_3}(v_i^0)} \frac{p_i(\Delta\theta_i)^{\alpha_1} (\Delta v_i)^{\alpha_2} (\Delta m_i)^{\alpha_3}}{\alpha_1! \alpha_2! \alpha_3!}, \end{aligned}$$

where $1 \leq i \leq k_0$ and $1 \leq j \leq 2r + 1$. Since $L_j^{\alpha_1, \alpha_2, \alpha_3} = 0$ as $\alpha_3 \geq 1$, we further obtain that

$$\beta_{ji}^{(r)} = \frac{2\Delta p_i}{\sigma_j^0} 1_{\{j=1\}} + \frac{1}{\sigma_i^0} \sum_{\alpha_1 + \alpha_2 \leq r} \frac{L_j^{\alpha_1, \alpha_2, 0}}{N_j^{\alpha_1, \alpha_2, 0}(v_i^0)} \frac{p_i(\Delta\theta_i)^{\alpha_1} (\Delta v_i)^{\alpha_2}}{\alpha_1! \alpha_2!}.$$

Therefore, $\beta_{ji}^{(r)}$ are polynomials of $\Delta p_i, \Delta\theta_i, \Delta v_i$, while $\gamma_j^{(r)}$ are polynomials of $\Delta\theta_i, \Delta v_i, \Delta m_i$, for $1 \leq i \leq k_0, 1 \leq j \leq 2r + 1$.

Suppose that there is a sequence of G tending to G_0 (in W_r distance) such that all coefficients of its r -minimal form in (29) vanish. It can be checked that $\beta_{ji}^{(r)}/W_r^r(G, G_0) \rightarrow 0$ for all even $j \in [1, 2r + 1]$

entails that $\Delta\theta_i/W_r^r(G, G_0) \rightarrow 0$ for all $1 \leq i \leq k_0$. Similarly, $\beta_{ji}^{(r)}/W_r^r(G, G_0) \rightarrow 0$ for all odd $j \in [3, 2r+1]$ entails that $\Delta v_i/W_r^r(G, G_0) \rightarrow 0$ for all $1 \leq i \leq k_0$. So, as $\beta_{1i}^{(r)}/W_r^r(G, G_0) \rightarrow 0$, we obtain $\Delta p_i/W_r^r(G, G_0) \rightarrow 0$. It follows that, as $\beta_{ji}^{(r)}/W_r^r(G, G_0) \rightarrow 0$ for all $1 \leq j \leq 2r+1$, we must have $\Delta p_i/W_r^r(G, G_0), \Delta\theta_i/W_r^r(G, G_0), \Delta v_i/W_r^r(G, G_0) \rightarrow 0$ for all $1 \leq i \leq k_0$. These results imply that

$$\frac{\sum_{i=1}^{k_0} |\Delta p_i| + p_i(|\Delta\theta_i|^r + |\Delta v_i|^r)}{W_r^r(G, G_0)} \rightarrow 0.$$

If $\Delta m_i = 0$ for all $1 \leq i \leq k_0$, then by means of Lemma 3.1, $\sum_{i=1}^{k_0} |\Delta p_i| + p_i(|\Delta\theta_i|^r + |\Delta v_i|^r) = D_r(G_0, G) \asymp W_r^r(G, G_0)$, which contradicts with the above limit. Therefore, we must have $\max_{1 \leq i \leq k_0} |\Delta m_i| > 0$.

Turning to $\gamma_l^{(r)}$ and the fact that $\Delta p_i/W_r^r(G, G_0), \Delta\theta_i/W_r^r(G, G_0), \Delta v_i/W_r^r(G, G_0) \rightarrow 0$, if $\gamma_l^{(r)}/W_r^r(G, G_0) \rightarrow 0$ as $1 \leq l \leq 2r$, we also have that

$$\left(\sum_{i=1}^{k_0} \sum_{\alpha_3 \leq r} \frac{U_l^{0,0,\alpha_3}(m_i^0) p_i(\Delta m_i)^{\alpha_3}}{V_l^{0,0,\alpha_3}(v_i^0) \alpha_3!} \right) / W_r^r(G, G_0) \rightarrow 0.$$

We can verify that as $1 \leq l \leq 2r$ is odd, $U_j^{0,0,\alpha_3}(m_i^0) = 0$ for all $\alpha_3 \leq r$ and $1 \leq i \leq k_0$. Additionally, as $1 \leq l \leq 2r$ is even, the above system of limits becomes

$$\left(\sum_{i_1-i_2=l/2} \frac{q_{i_1,i_2}}{i_1!} \sum_{i=1}^{k_0} \frac{p_i(m_i^0)^{i_1-2i_2-1} (\Delta m_i)^{i_1}}{(\sigma_i^0)^l} \right) / W_r^r(G_0, G) \rightarrow 0, \quad (30)$$

where $1 \leq i_1 \leq r, i_2 \leq (i_1-1)/2$ as i_1 is odd or $i_2 \leq i_1/2-1$ as i_1 is even. Here, $q_{i,j}$ are the integer coefficients that appear in the high order derivatives of $f(x|\theta, \sigma, m)$ with respect to m :

$$\frac{\partial^{s+1} f}{\partial m^{s+1}} = \left[\sum_{j=0}^{(s-1)/2} \frac{q_{(s+1),j} m^{s-2j}}{\sigma^{2s+2-2j}} (x-\theta)^{2s-2j+1} \right] f\left(\frac{x-\theta}{\sigma}\right) f\left(\frac{m(x-\theta)}{\sigma}\right)$$

when s is an odd number and

$$\frac{\partial^{s+1} f}{\partial m^{s+1}} = \left[\sum_{j=0}^{s/2} \frac{q_{(s+1),j} m^{s-2j}}{\sigma^{2s+2-2j}} (x-\theta)^{2s-2j+1} \right] f\left(\frac{x-\theta}{\sigma}\right) f\left(\frac{m(x-\theta)}{\sigma}\right)$$

when s is an even number. For instance, when $s = 0$, we have $q_{1,0} = 2$ and when $s = 1$, we have $q_{2,0} = -2$.

Summarizing, in order for all the coefficients in the r -minimal form (29) to vanish, i.e we have $\beta_{ji}^{(r)}, \gamma_l^{(r)}/W_r^r(G, G_0) \rightarrow 0$, the system of limits (30) is the key factor to determine the singularity structure of $G_0 \in \mathcal{S}_3$. We are going to explore the structure of this system of limits under the specific settings of $G_0 \in \mathcal{S}_3$.

7.1.1 Singularity level of $G_0 \in \mathcal{S}_{31}$

Recall from the above argument that, as we have $\beta_{ji}^{(r)}/W_r^r(G, G_0)$ when $1 \leq i, l \leq k_0$ and $1 \leq j \leq 2r + 1$, we obtain $\Delta\theta_i, \Delta v_i, \Delta p_i/W_r^r(G, G_0) \rightarrow 0$ for all $1 \leq i \leq k_0$. Combining with Lemma 3.1, it follows that

$$\sum_{i=1}^{k_0} p_i |\Delta m_i|^r / W_r^r(G_1, G) \not\rightarrow 0. \quad (31)$$

Since we have $\max_{1 \leq i \leq k_0} |\Delta m_i| > 0$, a combination of (30) and (31) leads to

$$\left(\sum_{i_1 - i_2 = l/2} \frac{q_{i_1, i_2}}{i_1!} \sum_{i=1}^{k_0} \frac{p_i (m_i^0)^{i_1 - 2i_2 - 1} (\Delta m_i)^{i_1}}{(\sigma_i^0)^l} \right) / \sum_{i=1}^{k_0} p_i |\Delta m_i|^r \rightarrow 0, \quad (32)$$

for any even l such that $1 \leq l \leq 2r$. Let $q_i = p_i/\sigma_i^0$, $t_i^0 = m_i^0/\sigma_i^0$, and $\Delta t_i = \Delta m_i/\sigma_i^0$ for all $1 \leq i \leq k_0$, then the above limits can be rewritten as

$$\left(\sum_{i=1}^{k_0} \sum_{i_1 - i_2 = l/2} \frac{q_{i_1, i_2}}{i_1!} q_i (t_i^0)^{i_1 - 2i_2 - 1} (\Delta t_i)^{i_1} \right) / \sum_{i=1}^{k_0} q_i |\Delta t_i|^r \rightarrow 0, \quad (33)$$

where in the summation of the above display, $1 \leq i_1 \leq r$, $i_2 \leq (i_1 - 1)/2$ as i_1 is odd, or $i_2 \leq i_1/2 - 1$ as i_1 is even and l is an even number ranging from 2 to $2r$. These are the limits of the ratio of two semipolynomial functions. The existence of these limits will be shown to entail the existence of zeros of a system of polynomial equations.

Greedy extraction of limiting polynomials As explained in the main text, it is generally difficult to obtain all polynomial limits of the system of rational semipolynomial functions given by (33). However, it is possible to obtain a subset of polynomial limits via a greedy method of extraction. We shall demonstrate this technique for the specific case $r = 3$, and then present a general result, not unlike what we have done in subsections 4.1 and 4.2 for o-mixtures. For $r = 3$, we only have three possible choices of l in (33), which are $l = 2, 4$ and 6 . As $l = 2$, we have $(i_1, i_2) = (1, 0)$. As $l = 4$, we obtain $(i_1, i_2) \in \{(2, 0), (3, 1)\}$. Finally, as $l = 6$, we get $(i_1, i_2) = (3, 0)$. Here, we can compute that $q_{1,0} = 2, q_{2,0} = -2, q_{3,1} = -2, q_{3,0} = 2$. Therefore, as $r = 3$, the system of limits (33) becomes

$$\begin{aligned} & \left(\sum_{i=1}^{k_0} q_i \Delta t_i \right) / \sum_{i=1}^{k_0} q_i |\Delta t_i|^3 \rightarrow 0, \\ & \left(\sum_{i=1}^{k_0} q_i t_i^0 (\Delta t_i)^2 + \frac{1}{3} q_i (\Delta t_i)^3 \right) / \sum_{i=1}^{k_0} q_i |\Delta t_i|^3 \rightarrow 0, \\ & \left(\sum_{i=1}^{k_0} q_i (t_i^0)^2 (\Delta t_i)^3 \right) / \sum_{i=1}^{k_0} q_i |\Delta t_i|^3 \rightarrow 0. \end{aligned} \quad (34)$$

Denote $|\Delta t_{k_0}| := \max_{1 \leq i \leq k_0} \{|\Delta t_i|\}$. In each of the limiting expressions in the above display, we shall divide both the numerator and denominator of the left hand side by $|\Delta t_{k_0}|^\alpha$, where α is the smallest degree that appears in one of the monomials in the numerator. Since $|\Delta t_i|/|\Delta t_{k_0}|$ is bounded, there

exist a subsequence according to which $\Delta t_i/|\Delta t_{k_0}|$ tends to a constant, say k_i , for each $i = 1, \dots, k_0$. Note that at least one of the k_i is non-zero. Moreover, we obtain the following equations in the limit

$$\sum_{i=1}^{k_0} q_i^0 k_i = 0, \quad \sum_{i=1}^{k_0} q_i^0 t_i^0 (k_i)^2 = 0, \quad \sum_{i=1}^{k_0} q_i^0 (t_i^0)^2 (k_i)^3 = 0.$$

Since $q_i^0 = p_i^0/\sigma_i^0$, $t_i^0 = m_i^0/\sigma_i^0$ for all $1 \leq i \leq k_0$, by rescaling k_i , the above system of polynomial equations can be rewritten as

$$\sum_{i=1}^{k_0} p_i^0 k_i = 0, \quad \sum_{i=1}^{k_0} p_i^0 m_i^0 (k_i)^2 = 0, \quad \sum_{i=1}^{k_0} p_i^0 (m_i^0)^2 (k_i)^3 = 0.$$

Now we shall apply the greedy extraction technique to the general system (33). This involves dividing both the numerator and the denominator of the left hand side in each equation of the system by $(\Delta t_{k_0})^{l/2}$ for any $2 \leq l \leq 2r$ and l is even. This leads to the existence of solution for the following system of polynomial equations

$$\sum_{i=1}^{k_0} p_i^0 (m_i^0)^{l/2-1} k_i^{l/2} = 0, \quad (35)$$

where the index l is even and $2 \leq l \leq 2r$. In this system, at least one of k_i is non-zero.

At this point, by a contrapositive argument we immediately deduces that if system of polynomial equations (35) does *not* have a valid solution for the k_i , one of which must be non-zero, then G_0 is *not* r -singular relative to \mathcal{E}_{k_0} . It follows that $\ell(G_0|\mathcal{E}_{k_0}) \leq r - 1$. This connection motivates a deeper investigation into the behavior of the system of real polynomial equations (35).

Behavior of system of limiting polynomial equations We proceed to study the solvability of the system of polynomial equations like (35). Consider two parameter sequences $\mathbf{a} = \{a_i\}_{i=1}^{k_0}$, $\mathbf{b} = \{b_i\}_{i=1}^{k_0}$ such that $a_i > 0, b_i \neq 0$ for all $1 \leq i \leq l$ and b_i are pairwise different. Additionally, there exists two indices $1 \leq i_1 \neq j_1 \leq l$ such that $b_{i_1} b_{j_1} < 0$. We can think of a_i as taking the role of p_i^0 and b_i the role of m_i^0 .

Define $\bar{s}(k_0, \mathbf{a}, \mathbf{b})$ to be the *minimum* value of $s \geq 1$ such that the following system of polynomial equations

$$\sum_{i=1}^{k_0} a_i b_i^u c_i^{u+1} = 0, \quad \text{for } u = 0, 1, \dots, s \quad (36)$$

does not admit any *non-trivial* solution, by which we require that at least one of c_i is non-zero. For example, if $s = 2$, and $k_0 = 2$, the above system of polynomial equations is

$$a_1 c_1 + a_2 c_2 = 0, \quad a_1 b_1 c_1^2 + a_2 b_2 c_2^2 = 0, \quad a_1 b_1^2 c_1^3 + a_2 b_2^2 c_2^3 = 0.$$

In general, it is difficult to determine the exact value of $\bar{s}(k_0, \mathbf{a}, \mathbf{b})$ since it depends on the specific values of parameter sequences \mathbf{a} and \mathbf{b} . However, it is possible to obtain some nontrivial bounds:

Proposition 7.1. *Let $k_0 \geq 2$.*

(a) *If for any subset I of $\{1, 2, \dots, k_0\}$ we have $\sum_{i \in I} a_i \prod_{j \in I \setminus \{i\}} b_j \neq 0$, then $\bar{s}(k_0, \mathbf{a}, \mathbf{b}) \leq k_0 - 1$.*

(b) If there is a subset I of $\{1, 2, \dots, k_0\}$ such that $\sum_{i \in I} a_i \prod_{j \in I \setminus \{i\}} b_j = 0$, then $\bar{s}(k_0, \mathbf{a}, \mathbf{b}) = \infty$.

(c) Under the same condition as that of part (a):

If $k_0 = 2$, then $\bar{s}(k_0, \mathbf{a}, \mathbf{b}) = 1$.

If $k_0 = 3$, and $\sum_{i=1}^{k_0} a_i \prod_{j \neq i, j \leq k_0} b_j > 0$, then $\bar{s}(k_0, \mathbf{a}, \mathbf{b}) = 1$. Otherwise, $\bar{s}(k_0, \mathbf{a}, \mathbf{b}) = 2$.

Remarks (i) Applying part (a) of this proposition to system (35), since $G_0 \in \mathcal{S}_{31}$, i.e. $P_3(\mathbf{p}^0, \boldsymbol{\eta}^0) = \sum_{i=1}^{k_0} p_i^0 \prod_{j \neq i} m_j^0 \neq 0$, G_0 is not $\bar{s}(k_0, \{p_i^0\}_{i=1}^{k_0}, \{m_i^0\}_{i=1}^{k_0}) + 1$ -singular relative to \mathcal{E}_{k_0} . Therefore, the singularity level of G_0 is at most $\bar{s}(k_0, \{p_i^0\}_{i=1}^{k_0}, \{m_i^0\}_{i=1}^{k_0})$. (ii) Part (a) provides a mild condition of parameter sequences \mathbf{a}, \mathbf{b} under which a nontrivial finite upper bound can be obtained. A closer investigation of the proof establishes that this bound is tight, i.e., there exists (\mathbf{a}, \mathbf{b}) such that $\bar{s}(k_0, \mathbf{a}, \mathbf{b}) = k_0 - 1$ holds. This motivates the definition of \mathcal{S}_{31} . (iii) Part (b) suggests the possibility of infinite level of singularity, even as k_0 is fixed. We will show that this happens when $G_0 \in \mathcal{S}_{33}$. (iv) Part (c) suggests that the singularity levels of G_0 may be different for different values of $(\mathbf{p}^0, \boldsymbol{\eta}^0)$ for the same k_0 .

General bounds for singularity level of $G_0 \in \mathcal{S}_{31}$ So far, we assume that G_0 has exactly one homologous set without C(1) singularity of size k_0 . Now, we suppose that G_0 has more than one nonconformant homologous set without C(1) singularity of components, and that there are no Gaussian components (i.e., $P_1(\boldsymbol{\eta}^0) = \prod_{j=1}^{k_0} m_j^0 \neq 0$). It can be observed that the singularity level of G_0 can be bounded in terms of a number of system of polynomial equations of the same form as Eq. (35), which are applied to *disjoint* subsets of nonconformant homologous components. The application to each subset yields a corresponding system of polynomial limits like (33). If none of such systems admit non-trivial solutions, then we are absolutely certain that their corresponding systems of limiting equations cannot hold. As a consequence, we obtain that $\ell(G_0 | \mathcal{E}_{k_0}) \leq \bar{s}(G_0)$, where

$$\bar{s}(G_0) := \max_I \bar{s}(|I|, \{p_i^0\}_{i \in I}, \{m_i^0\}_{i \in I}), \quad (37)$$

where the maximum is taken over all nonconformant homologous subsets I of components of G_0 .

If, on the other hand, G_0 has one or more Gaussian components, in addition to having some nonconformant homologous subsets, then by combining the argument presented in Section 5.1 with the foregoing argument, we deduce that the singularity level of G_0 is at most $\max\{2, \bar{s}(G_0)\}$. Summarizing, we have the following theorem regarding the upper bound of singularity levels of $G_0 \in \mathcal{S}_{31}$ whose rigorous proof is deferred to Appendix B.

Theorem 7.1. Suppose that $G_0 \in \mathcal{S}_{31}$.

(a) If $P_1(\boldsymbol{\eta}^0) \neq 0$, then $\ell(G_0 | \mathcal{E}_{k_0}) \leq \bar{s}(G_0) \leq k^* - 1 \leq k_0 - 1$.

(b) If $P_1(\boldsymbol{\eta}^0) = 0$, then $\ell(G_0 | \mathcal{E}_{k_0}) \leq \max\{2, \bar{s}(G_0)\} \leq \max\{2, k^* - 1\} \leq \max\{2, k_0 - 1\}$.

where k^* is the maximum length among all nonconformant homologous sets without C(1) singularity of G_0 .

Exact calculations in special cases Since our proof method was to extract only an (incomplete) subset of polynomial limits, we could only speak of upper bounds of the singularity level, not lower bounds in general. For some special cases of $G_0 \in \mathcal{S}_{31}$, with extra work we can determine the exact singularity level of G_0 . This is based on the specific value of k^* , which is defined to be the maximum length among all nonconformant homologous sets without C(1) singularity of G_0 in Theorem 7.1:

Proposition 7.2. (Exact singularity level) Assume that $G_0 \in \mathcal{S}_{31}$ and $P_1(\boldsymbol{\eta}^0) \neq 0$.

(a) If $k^* = 2$, then $\ell(G_0|\mathcal{E}_{k_0}) = 1$.

(b) Let $k^* = 3$. In addition, if all homologous sets I of G_0 such that $|I| = k^*$ satisfy $\sum_{i \in I} p_i^0 \prod_{j \in I \setminus \{i\}} m_j^0 > 0$, then $\ell(G_0|\mathcal{E}_{k_0}) = 1$. Otherwise, $\ell(G_0|\mathcal{E}_{k_0}) = 2$.

7.1.2 Singularity structure of \mathcal{S}_{32}

For the simplicity of the argument in this section, we go back to the simple setting of G_0 , i.e., G_0 has only one homologous set of size k_0 . Since $G_0 \in \mathcal{S}_{32}$, we have $P_3(\mathbf{p}^0, \boldsymbol{\eta}^0) = \sum_{i=1}^{k_0} p_i^0 \prod_{j \neq i} m_j^0 = 0$. This entails that $\bar{s}(k_0, \{p_i^0\}, \{m_i^0\}) = \infty$ according to part (b) of Proposition 7.1. As a result, $\bar{s}(G_0) = \infty$, i.e., the upper bound given by Theorem 7.1, that is, $\ell(G_0|\mathcal{E}_{k_0}) \leq \bar{s}(G_0)$, is no longer meaningful for \mathcal{S}_{32} . This does not necessarily imply that the singularity level for $G_0 \in \mathcal{S}_{32}$ is infinite. It simply means that the system of polynomial equations in (35) will not lead to any contradiction for any order r . In fact, these equations described by (35) are no longer sufficient to express the polynomial limits of the system (32). The issue is that our greedy extraction of polynomial limits for the system (32) treats each equation of the system separately. For instance, in system (34), a special case of system (32) when $r = 3$, we do not consider the interaction between two summations $\sum_{i=1}^{k_0} q_i t_i^0 (\Delta t_i)^2$ and $\sum_{i=1}^{k_0} \frac{1}{3} q_i (\Delta t_i)^3$ in the numerator of the second limit. As a result, the limiting polynomials obtained are dependent only on the lowest order monomial terms that appear in the numerator of each of the r -minimal form's coefficients.

To go further with \mathcal{S}_{32} , we introduce a more sophisticated technique for the polynomial limit extraction, which seeks to partially account for the interactions among different summations in the numerators of all the limits in system (32). This can be achieved by keeping not only the lowest order monomial in the numerator of the r -form's coefficient, but also the second lowest order monomials. As a result, we can extract a larger set of polynomial limits than (35). This would allow us to obtain a tighter bound of the singularity level for elements of \mathcal{S}_{32} . Although our extraction technique is general, the system of limiting polynomials that can be extracted is difficult to express explicitly for large values of k_0 . For this reason in the following we shall illustrate this technique of polynomial limit extraction on a specific case of $k_0 = 2$.

Proposition 7.3. Assume that $G_0 \in \mathcal{S}_{32}$ and G_0 has only one homologous set of size k_0 . Then as $k_0 = 2$, we have $\ell(G_0|\mathcal{E}_{k_0}) = 3$.

Remark: (i) The assumption that G_0 has only one homologous set is just for the convenience of the argument. The conclusion of this proposition still holds when $G_0 \in \mathcal{S}_{32}$ has multiple homologous sets and the maximum length of homologous sets with C(1) singularity is 2. (ii) By using the same technique, we can demonstrate that $\ell(G_0|\mathcal{E}_{k_0}) = k_0 + 1$ when $k_0 \leq 5$ and $G_0 \in \mathcal{S}_{32}$ has only one homologous set of size k_0 . We conjecture that this result also holds for general k_0 .

Proof. The proof proceeds in two main steps

Step 1: We will demonstrate that G_0 is 3-singular relative to \mathcal{E}_{k_0} . As $r = 3$, the system (32) consists of the following limiting equations, as $q_i \rightarrow q_i^0 > 0$ and $\Delta t_i \rightarrow 0$ for all $i = 1, 2$,

$$\begin{aligned} \sum_{i=1}^2 q_i \Delta t_i / \sum_{i=1}^2 q_i |\Delta t_i|^3 &\rightarrow 0, \\ \left(\sum_{i=1}^2 q_i t_i^0 (\Delta t_i)^2 + \frac{1}{3} q_i (\Delta t_i)^3 \right) / \sum_{i=1}^2 q_i |\Delta t_i|^3 &\rightarrow 0, \\ \left(\sum_{i=1}^2 q_i (t_i^0)^2 (\Delta t_i)^3 \right) / \sum_{i=1}^2 q_i |\Delta t_i|^3 &\rightarrow 0, \end{aligned}$$

where $q_i = p_i/\sigma_i^0$, $q_i^0 = p_i^0/\sigma_i^0$, $t_i^0 = m_i^0/\sigma_i^0$, and $\Delta t_i = \Delta m_i/\sigma_i^0$ for all $i = 1, 2$. The condition of $C(1)$ singularity means $P_3(\mathbf{p}^0, \mathbf{\eta}^0) = 0$. That is $p_1^0 m_2^0 + p_2^0 m_1^0 = 0$. So, $q_1^0 t_2^0 + q_2^0 t_1^0 = 0$. By choosing $\Delta t_2 = 1/n$, $\Delta t_1 = \frac{1}{n} \left(-\frac{q_2}{q_1} + \frac{1}{n^4} \right)$ where $q_1 = q_1^0 + 1/n$ and $q_2 = -q_1 t_2^0/t_1^0 + 1/n^2$, we can check that all of the above limits are satisfied. Hence, G_0 is 3-singular relative to \mathcal{E}_{k_0} .

Step 2: It remains to show that G_0 is *not* 4-singular relative to \mathcal{E}_{k_0} , and hence, G_0 's singularity level is 3. Let $r = 4$, the system (32) consists of the following limiting equations

$$\begin{aligned} \sum_{i=1}^2 q_i^n \Delta t_i^n / \sum_{i=1}^2 q_i^n |\Delta t_i^n|^4 &\rightarrow 0, \\ \left(\sum_{i=1}^2 q_i t_i^0 (\Delta t_i)^2 + \frac{1}{3} q_i (\Delta t_i)^3 \right) / \sum_{i=1}^2 q_i |\Delta t_i|^4 &\rightarrow 0, \\ \left(\sum_{i=1}^2 \frac{1}{3} q_i (t_i^0)^2 (\Delta t_i)^3 + \frac{1}{4} q_i t_i^0 (\Delta t_i)^4 \right) / \sum_{i=1}^2 q_i |\Delta t_i|^4 &\rightarrow 0, \\ \sum_{i=1}^2 q_i (t_i^0)^3 (\Delta t_i)^4 / \sum_{i=1}^2 q_i |\Delta t_i|^4 &\rightarrow 0. \end{aligned}$$

In order to account for the second-lowest order monomials of the numerator in each of the equations, we raise the order of the denominator in each equation to the former. That is,

$$\begin{aligned} K_1 &:= \sum_{i=1}^2 q_i \Delta t_i / \sum_{i=1}^2 q_i |\Delta t_i|^2 \rightarrow 0, \\ K_2 &:= \left(\sum_{i=1}^2 q_i t_i^0 (\Delta t_i)^2 + \frac{1}{3} q_i (\Delta t_i)^3 \right) / \sum_{i=1}^2 q_i |\Delta t_i|^3 \rightarrow 0, \\ K_3 &:= \left(\sum_{i=1}^2 \frac{1}{3} q_i (t_i^0)^2 (\Delta t_i)^3 + \frac{1}{4} q_i t_i^0 (\Delta t_i)^4 \right) / \sum_{i=1}^2 q_i |\Delta t_i|^4 \rightarrow 0, \\ K_4 &:= \sum_{i=1}^2 q_i (t_i^0)^3 (\Delta t_i)^4 / \sum_{i=1}^2 q_i |\Delta t_i|^4 \rightarrow 0. \end{aligned}$$

We assume without loss of generality that $|\Delta t_2|$ is the maximum between $|\Delta t_1|$ and $|\Delta t_2|$. Denote $\Delta t_1 = k_1 \Delta t_2$ where $k_1 \in [-1, 1]$ and $k_1 \rightarrow k_1'$. The vanishing of K_1 yields $q_1^0 k_1' + q_2^0 = 0$. So, $k_1' = -q_2^0/q_1^0 = t_2^0/t_1^0$.

Divide both the numerator and denominator of K_1 by $(\Delta t_2)^2$, we obtain $(q_1 k_1 + q_2)/\Delta t_2 \rightarrow 0$. Write $u = k_1 + q_2/q_1$, then $q_1 u/\Delta t_2 \rightarrow 0$, which implies that $u/\Delta t_2 \rightarrow 0$.

Next, divide both the numerator and denominator of K_2 by $(\Delta t_2)^3$, we obtain

$$\left(\sum_{i=1}^2 q_i t_i^0 (\Delta t_i)^2 + \frac{1}{3} q_i (\Delta t_i)^3 \right) / (\Delta t_2)^3 \rightarrow 0.$$

Plug in the formula of k_1 and the fact that $u/\Delta t_2 \rightarrow 0$, it follows that

$$\left(q_1 t_1^0 \left(\frac{q_2}{q_1} \right)^2 + q_2 t_2^0 \right) / (\Delta t_2) \rightarrow -\frac{1}{3} (q_1^0 (k_1')^3 + q_2^0).$$

Thus, we get $P_1 := (t_1^0 q_2 + t_2^0 q_1)/\Delta t_2 \rightarrow -\frac{q_1^0}{3q_2^0} (q_1^0 (k_1')^3 + q_2^0)$. It is simple to verify that this limit is non-zero, otherwise we would have $q_1^0 = q_2^0$, which violates the definition that G_0 does not have C(2) singularity, i.e $G_0 \in \mathcal{S}_{32}$.

Continuing, divide both the numerator and denominator of K_3 by $(\Delta t_2)^4$, and with the same argument, we obtain $P_2 := (t_1^0 q_2 - t_2^0 q_1)(t_1^0 q_2 + t_2^0 q_1)/\Delta t_2 \rightarrow -\frac{3(q_1^0)^2}{4q_2^0} (q_1^0 t_1^0 (k_1')^4 + q_2^0 t_2^0)$.

By dividing P_2 by P_1 and let it to vanish, we can extract the following polynomial in the limit:

$$4(q_1^0 (k_1')^3 + q_2^0)(t_1^0 q_2^0 - t_2^0 q_1^0) = 9q_1^0 (q_1^0 t_1^0 (k_1')^4 + q_2^0 t_2^0).$$

By plugging in $k_1' = -q_2^0/q_1^0$ and $t_1^0 q_2^0 + t_2^0 q_1^0 = 0$, we can deduce that $q_1^0 = q_2^0$, which is a contradiction. Thus, we conclude that G_0 is not 4-singular relative to \mathcal{E}_{k_0} . \square

7.1.3 Singularity level of $G_0 \in \mathcal{S}_{33}$

As we can see from the proof of Proposition 7.2, the condition of without C(2) singularity plays a major role in guaranteeing that $G_0 \in \mathcal{S}_{32}$ is not 4-singular relative to \mathcal{E}_{k_0} when G_0 has only one homologous set of $k_0 = 2$. Therefore, for elements G_0 in \mathcal{S}_{33} , we expect the singularity level of G_0 may be very large. In fact, we can show that

Theorem 7.2. *If $G_0 \in \mathcal{S}_{33}$, then $\ell(G_0|\mathcal{E}_{k_0}) = \infty$.*

Proof. Here, we present the proof for $k_0 = 2$. For general values of k_0 , the proof is similar and deferred to Appendix B. For $k_0 = 2$, the condition that $G_0 \in \mathcal{S}_{33}$ entails $P_4(\mathbf{p}^0, \boldsymbol{\eta}^0) = 0$, i.e $p_1^0/\sigma_1^0 = p_2^0/\sigma_2^0$ and $m_1^0/\sigma_1^0 = -m_2^0/\sigma_2^0$. By choosing $\Delta m_1/\sigma_1^0 = -\Delta m_2/\sigma_2^0$, $p_1 = p_2 = p_1^0 = p_2^0$, we can check that

$$\sum_{i=1}^2 \frac{p_i (m_i^0)^u (\Delta m_j)^v}{(\sigma_i^0)^{u+v+1}} = 0,$$

for all odd numbers $u \in [1, v]$ when v is even number, or for all even numbers $u \in [0, v]$ when v is odd number.

Take order $r \geq 1$ to be an arbitrary natural number. Incorporating the identity in the previous display into (29) and (30), we obtain the vanishing of all $\gamma_l^{(r)}/W_r^r(G_1, G)$ for all $1 \leq l \leq 2r$ and l is even. If we choose $\Delta \theta_i = \Delta v_i = 0$ for all $1 \leq i \leq 2$, we also have the coefficients $\beta_{ji}^{(r)}/W_r^r(G, G_0) = 0$ for all $1 \leq i \leq 2$ and $1 \leq j \leq 2r + 1$. Additionally, we also have $\gamma_l^{(r)}/W_r^r(G, G_0) = 0$ for all $1 \leq l \leq 2r$ and l is odd. Hence, G_0 is r -singular relative to \mathcal{E}_{k_0} for any $r \geq 1$. As a consequence, $\ell(G_0|\mathcal{E}_{k_0}) = \infty$. \square

8 Appendix B

This Appendix contains the remaining proofs of the results presented in the paper.

8.1 Proofs for Section 3

PROOF OF THEOREM 3.1 Since the proofs for part (iii) and (iv) are similar, we only provide the proof for part (iii). The proof of this part is the generalization of that of part (c) in Theorem 3.2 in [Ho and Nguyen, 2016b]. By means of Taylor expansion up to r -th order, we have

$$\begin{aligned} h^2(p_G, p_{G_0}) &< \int_{x \in \mathcal{X}} \frac{(p_G(x) - p_{G_0}(x))^2}{p_{G_0}(x)} dx = \int_{x \in \mathcal{X}} \frac{\left(\sum_{l=1}^{T_r} \xi_l^{(r)}(G) H_l^{(r)}(x) + R_r(x) \right)^2}{p_{G_0}(x)} dx \\ &= \int_{x \in \mathcal{X}} \frac{R_r^2(x)}{p_{G_0}(x)} dx. \end{aligned}$$

Here, $R_r(x)$ has the following form

$$R_r(x) = \sum_{i=1}^{k_0} \sum_{j=1}^{s_i} \sum_{|\alpha|=r+1} \frac{r+1}{\alpha!} (\Delta \eta_{ij})^\alpha \int_0^1 (1-t)^r \frac{\partial^{r+1} f}{\partial \eta^\alpha} (x | \eta_i^0 + t \Delta \eta_{ij}) dt.$$

Hence, as $p_{G_0}(x) > p_i^0 f(x | \eta_i^0)$ for all $1 \leq i \leq k_0$, for any $s < r+1$, we have

$$\begin{aligned} \frac{h^2(p_G, p_{G_0})}{W_1^{2s}(G, G_0)} &< \int_{x \in \mathcal{X}} \frac{R_r^2(x)}{W_1^{2s}(G, G_0) p_{G_0}(x)} dx \\ &< \sum_{i=1}^{k_0} \int_{x \in \mathcal{X}} \frac{\left(\sum_{j=1}^{s_i} \sum_{|\alpha|=r+1} \frac{r+1}{\alpha!} (\Delta \eta_{ij})^\alpha \int_0^1 (1-t)^r \frac{\partial^{r+1} f}{\partial \eta^\alpha} (x | \eta_i^0 + t \Delta \eta_{ij}) dt \right)^2}{W_1^{2s}(G, G_0) p_i^0 f(x | \eta_i^0)} dx \\ &\lesssim \sum_{i=1}^{k_0} \int_{x \in \mathcal{X}} \frac{\sum_{j=1}^{s_i} \sum_{|\alpha|=r+1} \left(\frac{r+1}{\alpha!} (\Delta \eta_{ij})^\alpha \int_0^1 (1-t)^r \frac{\partial^{r+1} f}{\partial \eta^\alpha} (x | \eta_i^0 + t \Delta \eta_{ij}) dt \right)^2}{W_1^{2s}(G, G_0) p_i^0 f(x | \eta_i^0)} dx, \end{aligned}$$

where the last inequality comes from Cauchy-Schwarz's inequality. Now, for any $s < r+1$, by utilizing Lemma 3.1, we obtain

$$\frac{|(\Delta \eta_{ij})^\alpha|}{W_1^s(G, G_0)} \asymp \frac{|(\Delta \eta_{ij})^\alpha|}{D_1^s(G_0, G)} < \frac{|(\Delta \eta_{ij})^\alpha|}{\|\Delta \eta_{ij}\|^s} \rightarrow 0, \quad (38)$$

for any $|\alpha| = r+1$. According to the hypothesis, as $\Delta \eta_{ij} < c_0$, we have

$$\int_{x \in \mathcal{X}} \frac{\left(\int_0^1 (1-t)^r \frac{\partial^{r+1} f}{\partial \eta^\alpha} (x | \eta_i^0 + t \Delta \eta_{ij}) dt \right)^2}{p_i^0 f(x | \eta_i^0)} dx < \int_{x \in \mathcal{X}} \frac{\left(\frac{\partial^{r+1} f}{\partial \eta^\alpha} (x | \eta_i^0 + t \Delta \eta_{ij}) \right)^2}{p_i^0 f(x | \eta_i^0)} dx < \infty. \quad (39)$$

Combining (38) and (39), we achieve $h(p_G, p_{G_0})/W_1^s(G, G_0) \rightarrow 0$, which yields the conclusion of this part.

8.2 Proofs for Section 4

PROOF OF LEMMA 4.1 For any $k_0 \geq 1$ and k_0 different pairs $\eta_1 = (\theta_1, \sigma_1, m_1), \dots, \eta_{k_0} = (\theta_{k_0}, \sigma_{k_0}, m_{k_0})$, let $\alpha_{ij} \in \mathbb{R}$ for $i = 1, \dots, 4, j = 1, \dots, k_0$ such that for almost all $x \in \mathbb{R}$

$$\sum_{j=1}^{k_0} \alpha_{1j} f(x|\eta_j) + \alpha_{2j} \frac{\partial f}{\partial \theta}(x|\eta_j) + \alpha_{3j} \frac{\partial f}{\partial \sigma^2}(x|\eta_j) \alpha_{4j} \frac{\partial f}{\partial m}(x|\eta_j) = 0.$$

We can rewrite the above equation as

$$\sum_{j=1}^{k_0} \left\{ [\beta_{1j} + \beta_{2j}(x - \theta_j) + \beta_{3j}(x - \theta_j)^2] \Phi\left(\frac{m_j(x - \theta_j)}{\sigma_j}\right) \exp\left(-\frac{(x - \theta_j)^2}{2\sigma_j^2}\right) + (\gamma_{1j} + \gamma_{2j}(x - \theta_j)) f\left(\frac{m_j(x - \theta_j)}{\sigma_j}\right) \exp\left(-\frac{(x - \theta_j)^2}{2\sigma_j^2}\right) \right\} = 0, \quad (40)$$

where $\beta_{1j} = \frac{2\alpha_{1j}}{\sqrt{2\pi}\sigma_j} - \frac{\alpha_{3j}}{\sqrt{2\pi}\sigma_j^3}$, $\beta_{2j} = \frac{2\alpha_{2j}}{\sqrt{2\pi}\sigma_j^3}$, $\beta_{3j} = \frac{\alpha_{3j}}{\sqrt{2\pi}\sigma_j^5}$, $\gamma_{1j} = -\frac{2\alpha_{2j}m_j}{\sqrt{2\pi}\sigma_j^2}$, and $\gamma_{2j} = -\frac{\alpha_{3j}m_j}{\sqrt{2\pi}\sigma_j^4} + \frac{2\alpha_{4j}}{\sqrt{2\pi}\sigma_j^2}$ for all $j = 1, \dots, k_0$.

”Only if” direction: Assume by contrary that the conclusion does not hold, i.e., both type A and type B conditions do not hold. Denote $\sigma_{j+k_0} = \frac{\sigma_j^2}{1+m_j^2}$ for all $1 \leq j \leq k_0$. For the simplicity of the argument, we assume that σ_i are pairwise different and $\frac{\sigma_i^2}{1+m_i^2} \notin \left\{ \sigma_j^2 : 1 \leq j \leq k_0 \right\}$ for all $1 \leq i \leq k_0$. The argument for the other cases is similar. Now, σ_j are pairwise different as $1 \leq j \leq 2k_0$. The equation (40) can be rewritten as

$$\sum_{j=1}^{2k_0} \left\{ [\beta_{1j} + \beta_{2j}(x - \theta_j) + \beta_{3j}(x - \theta_j)^2] \Phi\left(\frac{m_j(x - \theta_j)}{\sigma_j}\right) \exp\left(-\frac{(x - \theta_j)^2}{2\sigma_j^2}\right) \right\} = 0, \quad (41)$$

where $m_j = 0$, $\theta_{j+k_0} = \theta_j$, $\beta_{1(j+k_0)} = \frac{2\gamma_{1j}}{\sqrt{2\pi}}$, $\beta_{2(j+k_0)} = \frac{2\gamma_{2j}}{\sqrt{2\pi}}$, $\beta_{3j} = 0$ as $k_0 + 1 \leq j \leq 2k_0$.

Denote $\bar{i} = \arg \max_{1 \leq i \leq 2k_0} \{\sigma_i\}$. Multiply both sides of (41) with $\exp\left(\frac{(x - \theta_{\bar{i}})^2}{2\sigma_{\bar{i}}^2}\right) / \Phi\left(\frac{m_{\bar{i}}(x - \theta_{\bar{i}})}{\sigma_{\bar{i}}}\right)$ and let $x \rightarrow +\infty$ if $m_{\bar{i}} \geq 0$ or let $x \rightarrow -\infty$ if $m_{\bar{i}} < 0$ on both sides of the new equation, we obtain $\beta_{1\bar{i}} + \beta_{2\bar{i}}(x - \theta_{\bar{i}}) + \beta_{3\bar{i}}(x - \theta_{\bar{i}})^2 \rightarrow 0$. It implies that $\beta_{1\bar{i}} = \beta_{2\bar{i}} = \beta_{3\bar{i}} = 0$. Repeatedly apply the same argument to the remaining σ_i until we obtain $\beta_{1i} = \beta_{2i} = \beta_{3i} = 0$ for all $1 \leq i \leq 2k_0$. It is equivalent to $\alpha_{1i} = \alpha_{2i} = \alpha_{3i} = \alpha_{4i} = 0$ for all $1 \leq i \leq k_0$, which is a contradiction.

”If” direction: There are two possible scenarios.

Type A singularity There exists some $m_j = 0$ as $1 \leq j \leq k_0$. In this case, we assume that $m_1 = 0$. If we choose $\alpha_{1j} = \alpha_{2j} = \alpha_{3j} = \alpha_{4j} = 0$ for all $2 \leq j \leq k_0$, then equation (40) can be rewritten as

$$\frac{\beta_{11}}{2} + \frac{\gamma_{11}}{\sqrt{2\pi}} + \left(\frac{\beta_{21}}{2} + \frac{\gamma_{21}}{\sqrt{2\pi}} \right) (x - \theta_1) + \frac{\beta_{31}}{2} (x - \theta_1)^2 = 0.$$

By choosing $\alpha_{31} = 0$, $\alpha_{11} = \frac{\alpha_{21}m_1}{\sqrt{2\pi}\sigma_1}$, $\alpha_{21} = -\frac{\alpha_{41}\sigma_1}{\sqrt{2\pi}}$, the above equation always equal to 0. Since $\alpha_{11}, \alpha_{21}, \alpha_{41}$ are not necessarily zero, the first-order identifiability (i.e., linear independence condition) is violated.

Type B singularity There exists indices $1 \leq i \neq j \leq k_0$ such that $\left(\frac{\sigma_i^2}{1+m_i^2}, \theta_i\right) = \left(\frac{\sigma_j^2}{1+m_j^2}, \theta_j\right)$. Without loss of generality, we assume that $i = 1, j = 2$. If we choose $\alpha_{1j} = \alpha_{2j} = \alpha_{3j} = \alpha_{4j} = 0$ for all $3 \leq j \leq k_0$, then equation in (40) can be rewritten as

$$\sum_{j=1}^2 \left\{ [\beta_{1j} + \beta_{2j}(x - \theta_j) + \beta_{3j}(x - \theta_j)^2] \Phi\left(\frac{m_j(x - \theta_j)}{\sigma_j}\right) \exp\left(-\frac{(x - \theta_j)^2}{2\sigma_j^2}\right) \right\} + \frac{1}{\sqrt{2\pi}} \left(\sum_{j=1}^2 \gamma_{1j} + \sum_{j=1}^2 \gamma_{2j}(x - \theta_1) \right) \exp\left(-\frac{(m_1^2 + 1)(x - \theta_1)^2}{2\sigma_1^2}\right) = 0.$$

Now, we choose $\alpha_{1j} = \alpha_{2j} = \alpha_{3j} = 0$ for all $1 \leq j \leq 2$, $\frac{\alpha_{41}}{\sigma_1^2} + \frac{\alpha_{42}}{\sigma_2^2} = 0$ then the above equation always hold. Since α_{41} and α_{42} need not be zero, the first-order identifiability condition is violated. This concludes the proof.

PROOF OF LEMMA 4.2 The proof proceeds via induction on $|\alpha|$. As $|\alpha| \leq 2$, we can easily check the conclusion of the lemma. Assume that the conclusion holds for any $|\alpha| \leq k - 1$. We shall demonstrate that it also holds for $|\alpha| = k$. Indeed, there are two settings:

Case 1: $\alpha_1 = k$ Under this setting, $\alpha_2 = \alpha_3 = 0$. From the induction hypothesis,

$$\begin{aligned} \frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} \partial v^{\alpha_2} \partial m^{\alpha_3}} &= \frac{\partial}{\partial \theta} \left(\frac{\partial^{|\alpha|-1} f}{\partial \theta^{\alpha_1-1} \partial v^{\alpha_2} \partial m^{\alpha_3}} \right) \\ &= \frac{\partial}{\partial \theta} \left(\sum_{\kappa \in \mathcal{F}_{|\alpha|-1}} \frac{P_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2} \partial m^{\kappa_3}} \right) \\ &= \sum_{\kappa \in \mathcal{F}_{|\alpha|-1}} \frac{P_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1+1} \partial v^{\kappa_2} \partial m^{\kappa_3}}, \\ &= \sum_{\kappa \in \mathcal{F}_{k-1}: \kappa_1=0} \frac{P_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1+1} \partial v^{\kappa_2} \partial m^{\kappa_3}} \\ &\quad + \sum_{\kappa \in \mathcal{F}_{k-1}: \kappa_1=1} \frac{P_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1-1, \alpha_2, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1+1} \partial v^{\kappa_2} \partial m^{\kappa_3}} \end{aligned} \quad (42)$$

where the second equality is due to the application of the hypothesis for $\alpha_1 - 1 + \alpha_2 + \alpha_3 = k - 1$. For any $\kappa \in \mathcal{F}_{k-1}$ such that $\kappa_1 = 1$,

$$\begin{aligned} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1+1} \partial v^{\kappa_2} \partial m^{\kappa_3}} &= \frac{\partial^{|\kappa|-1} f}{\partial v^{\kappa_2} \partial m^{\kappa_3}} \left(2 \frac{\partial f}{\partial v} - \frac{m^3 + m}{v} \frac{\partial f}{\partial m} \right) \\ &= 2 \frac{\partial^{|\kappa|} f}{\partial v^{\kappa_2+1} \partial m^{\kappa_3}} - \frac{\partial^{|\kappa|-1} f}{\partial v^{\kappa_2} \partial m^{\kappa_3}} \left(\frac{m^3 + m}{v} \frac{\partial f}{\partial m} \right). \end{aligned} \quad (43)$$

From the inductive hypothesis, since $|\kappa| = \kappa_2 + \kappa_3 + 1 \leq k - 1$,

$$\frac{\partial^{|\kappa|} f}{\partial v^{\kappa_2+1} \partial m^{\kappa_3}} = \sum_{\kappa' \in \mathcal{F}_{|\kappa|}} \frac{P_{0, \kappa_2+1, \kappa_3}^{\kappa'_1, \kappa'_2, \kappa'_3}(m)}{H_{0, \kappa_2+1, \kappa_3}^{\kappa'_1, \kappa'_2, \kappa'_3}(m) Q_{0, \kappa_2+1, \kappa_3}^{\kappa'_1, \kappa'_2, \kappa'_3}(v)} \frac{\partial^{|\kappa'|} f}{\partial \theta^{\kappa'_1} \partial v^{\kappa'_2} \partial m^{\kappa'_3}}. \quad (44)$$

In addition,

$$\frac{\partial^{|\kappa|-1} f}{\partial v^{\kappa_2} \partial m^{\kappa_3}} \left(\frac{m^3 + m}{v} \frac{\partial f}{\partial m} \right) = \sum_{\beta: |\beta| \leq |\kappa|, \beta_1 \leq \kappa_2, \beta_2 \leq \kappa_3+1} \frac{A_{\beta_1, \beta_2}(m)}{B_{\beta_1, \beta_2}(v)} \frac{\partial^{|\beta|} f}{\partial v^{\beta_1} \partial m^{\beta_2}}. \quad (45)$$

Since $|\beta| \leq |\kappa| \leq k - 1$, from the hypothesis,

$$\frac{\partial^{|\beta|} f}{\partial v^{\beta_1} \partial m^{\beta_2}} = \sum_{\kappa'' \in \mathcal{F}_{|\beta|}} \frac{P_{0, \beta_1, \beta_2}^{\kappa''_1, \kappa''_2, \kappa''_3}(m)}{H_{0, \beta_1, \beta_2}^{\kappa''_1, \kappa''_2, \kappa''_3}(m) Q_{0, \beta_1, \beta_2}^{\kappa''_1, \kappa''_2, \kappa''_3}(v)} \frac{\partial^{|\kappa''|} f}{\partial \theta^{\kappa''_1} \partial v^{\kappa''_2} \partial m^{\kappa''_3}}. \quad (46)$$

Combining equations (42), (43), (44), (45), and (46), we arrive at the conclusion of the lemma.

Case 2: $\alpha_1 \leq k - 1$ Under this setting, assume without loss of generality that $\alpha_2 \geq 1$.

$$\begin{aligned} \frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} \partial v^{\alpha_2} \partial m^{\alpha_3}} &= \frac{\partial}{\partial v} \left(\frac{\partial^{|\alpha|-1} f}{\partial \theta^{\alpha_1} \partial v^{\alpha_2-1} \partial m^{\alpha_3}} \right) \\ &= \frac{\partial}{\partial v} \left(\sum_{\kappa \in \mathcal{F}_{|\alpha|-1}} \frac{P_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2} \partial m^{\kappa_3}} \right) \\ &= \sum_{\kappa \in \mathcal{F}_{|\alpha|-1}} \frac{\partial}{\partial v} \left(\frac{P_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \right) \frac{\partial^{|\kappa|} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2} \partial m^{\kappa_3}} \\ &\quad + \sum_{\kappa \in \mathcal{F}_{|\alpha|-1}} \frac{P_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2+1} \partial m^{\kappa_3}}. \end{aligned} \quad (47)$$

Denote $A := \sum_{\kappa \in \mathcal{F}_{|\alpha|-1}} \frac{P_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2+1} \partial m^{\kappa_3}}$, we further have that

$$\begin{aligned} A &= \sum_{\kappa \in \mathcal{F}_{|\alpha|-1}: \kappa_3=0} \frac{P_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2+1} \partial m^{\kappa_3}} \\ &\quad + \sum_{\kappa \in \mathcal{F}_{|\alpha|-1}: \kappa_2=0, \kappa_3 \geq 1} \frac{P_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m)}{H_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(m) Q_{\alpha_1, \alpha_2-1, \alpha_3}^{\kappa_1, \kappa_2, \kappa_3}(v)} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2+1} \partial m^{\kappa_3}}. \end{aligned} \quad (48)$$

Since $m \neq 0$, for any $\kappa \in \mathcal{F}_{|\alpha|-1}$ such that $\kappa_2 = 0$ and $\kappa_3 \geq 1$, we have

$$\begin{aligned} \frac{\partial^{|\kappa|+1} f}{\partial \theta^{\kappa_1} \partial v^{\kappa_2+1} \partial m^{\kappa_3}} &= \frac{\partial^{|\kappa|-1} f}{\partial \theta^{\kappa_1} \partial m^{\kappa_3-1}} \left(-\frac{1}{v} \frac{\partial f}{\partial m} - \frac{m^2 + 1}{2mv} \frac{\partial^2 f}{\partial m^2} \right) \\ &= -\frac{1}{v} \frac{\partial^{|\kappa|} f}{\partial \theta^{\kappa_1} \partial m^{\kappa_3}} - \frac{\partial^{|\kappa|-1} f}{\partial \theta^{\kappa_1} \partial m^{\kappa_3-1}} \left(\frac{m^2 + 1}{2mv} \frac{\partial^2 f}{\partial m^2} \right). \end{aligned} \quad (49)$$

Since $|\kappa| = \kappa_1 + \kappa_3 \leq k - 1$ and $\kappa_1 \leq 1$, we have $(\kappa_1, 0, \kappa_3) \in \mathcal{F}_k$. Additionally, we can represent

$$\frac{\partial^{|\kappa|-1} f}{\partial \theta^{\kappa_1} \partial m^{\kappa_3-1}} \left(\frac{m^2 + 1}{2mv} \frac{\partial^2 f}{\partial m^2} \right) = \sum_{1 \leq \tau \leq \kappa_3+1} \frac{A'_\tau(m)}{B'_\tau(m)C'_\tau(v)} \frac{\partial^{\kappa_1+\tau} f}{\partial \theta^{\kappa_1} \partial m^\tau},$$

where $A'_\tau(m)$, $B'_\tau(m)$, $C'_\tau(v)$ are some polynomials of m and v . Since $\kappa_1 + \tau \leq \kappa_1 + \kappa_3 + 1 \leq k$ and $\kappa_1 \leq 1$, we have $(\kappa_1, 0, \tau) \in \mathcal{F}_k$. Combining these results with equations (47), (48), and (49), we achieve the conclusion of the lemma.

PROOF OF LEMMA 4.3 The proof of this lemma proceeds by induction on r . If $r = 1$,

$$\left\{ \frac{\partial^{|\alpha|} f}{\theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}} : (\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_r \right\} = \left\{ \frac{\partial f}{\partial \theta}, \frac{\partial f}{\partial v}, \frac{\partial f}{\partial m} \right\},$$

which are linearly independent with respect to $G_0 \in \mathcal{S}_0$ due to the conclusion of Lemma 4.1. Assume that the conclusion of the lemma holds up to r . We will demonstrate that it continues to hold for $r + 1$. In fact,

$$\left\{ \frac{\partial^{|\alpha|} f}{\theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}} : (\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_{r+1} \right\} = \left\{ \frac{\partial^{|\alpha|} f}{\theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}} : (\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_r \right\} \cup \left\{ \frac{\partial^{r+1} f}{\partial \theta \partial v^r}, \frac{\partial^{r+1} f}{\partial v^{r+1}}, \frac{\partial^{r+1} f}{\partial \theta \partial m^r}, \frac{\partial^{r+1} f}{\partial m^{r+1}} \right\}. \quad (50)$$

Assume that there are coefficients $\beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)}$ where $1 \leq i \leq k_0$ and $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_{r+1}$ such that for all x

$$\sum_{i=1}^{k_0} \sum_{(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_{r+1}} \beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)} \frac{\partial^{|\alpha|} f}{\theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0) = 0.$$

Using the fact from (50), we rewrite the above equation as

$$\begin{aligned} & \sum_{i=1}^{k_0} \sum_{(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_r} \beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)} \frac{\partial^{|\alpha|} f}{\theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0) + \beta_{1, r, 0}^{(i)} \frac{\partial^{r+1} f}{\partial \theta \partial v^r}(x|\eta_i^0) + \\ & \beta_{0, r+1, 0}^{(i)} \frac{\partial^{r+1} f}{\partial v^{r+1}}(x|\eta_i^0) + \beta_{1, 0, r}^{(i)} \frac{\partial^{r+1} f}{\partial \theta \partial m^r}(x|\eta_i^0) + \beta_{0, 0, r+1}^{(i)} \frac{\partial^{r+1} f}{\partial m^{r+1}}(x|\eta_i^0) = 0. \end{aligned} \quad (51)$$

Equation (51) can be rewritten as

$$\begin{aligned} & \sum_{i=1}^{k_0} \left(\sum_{j=1}^{2r+3} \gamma_{j,i}^{(r+1)} (x - \theta_i^0)^{j-1} \right) f \left(\frac{x - \theta_i^0}{\sigma_i^0} \right) \Phi \left(\frac{m_i^0 (x - \theta_i^0)}{\sigma_i^0} \right) \\ & + \sum_{i=1}^{k_0} \left(\sum_{j=1}^{2r+2} \tau_{j,i}^{(r+1)} (x - \theta_i^0)^{j-1} \right) \exp \left(-\frac{(m_i^0)^2 + 1}{2v_i^0} (x - \theta_i^0)^2 \right) = 0, \end{aligned}$$

where $\gamma_{j,i}^{(r+1)}$ are a combination of $\beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)}$ when $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_{r+1}$ and $\alpha_3 = 0$. Additionally, $\tau_{j,i}^{(r+1)}$ are a combination of $\beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)}$ when $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_{r+1}$. Due to the fact that there are no type

A or type B singularities in $\left\{\eta_1^0, \dots, \eta_{k_0}^0\right\}$, by using the same argument as that of the proof of Lemma 4.1, we obtain that $\gamma_{j,i}^{(r+1)} = 0$ for all $1 \leq i \leq k_0$, $1 \leq j \leq 2r+3$ and $\tau_{j,i}^{(r+1)} = 0$ for all $1 \leq i \leq k_0$, $1 \leq j \leq 2r+2$. It can be checked that $\gamma_{2r+3,i}^{(r+1)} = 0$ implies $\beta_{0,r+1,0}^{(i)} = 0$ while $\gamma_{2r+2,i}^{(r+1)} = 0$ implies $\beta_{1,r,0}^{(i)} = 0$ for all $1 \leq i \leq k_0$. Similarly, $\tau_{2r+2,i}^{(r+1)} = 0$ implies $\beta_{0,0,r+1}^{(i)} = 0$ while $\tau_{2r+1,i}^{(r+1)} = 0$ implies $\beta_{1,0,r}^{(i)} = 0$ for all $1 \leq i \leq k_0$. As a consequence, Eq. (51) is reduced to

$$\sum_{i=1}^{k_0} \sum_{(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_r} \beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)} \frac{\partial^{|\alpha|} f}{\theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0) = 0. \quad (52)$$

According to the hypothesis with r , we obtain that $\beta_{\alpha_1, \alpha_2, \alpha_3}^{(i)} = 0$ for all $1 \leq i \leq k_0$, $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{F}_r$. This concludes our proof.

PROOF OF PROPOSITION 4.1 From the formation of system of polynomial equations (21), if we choose $\beta_3 = 0$ (i.e., we only reduce to derivatives with respect to the location and scale parameter), then we have $P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m)/H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m)Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v) = 2^{\alpha_2}$ when $\alpha_3 = 0$ and $P_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m)/H_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(m)Q_{\alpha_1, \alpha_2, \alpha_3}^{\beta_1, \beta_2, \beta_3}(v) = 0$ as $\alpha_3 \geq 1$ for any v, m and $\alpha_1 + 2\alpha_2 + 2\alpha_3 = \beta_1 + 2\beta_2 + 2\beta_3$. This shows that the system of polynomial equations (21) contains the following system of equations

$$\sum_{j=1}^l \sum_{\alpha_1 + 2\alpha_2 = \beta_1 + 2\beta_2} \frac{2^{\alpha_2} d_j^2 a_j^{\alpha_1} b_j^{\alpha_2}}{\alpha_1! \alpha_2!} = 0, \quad (53)$$

where $\beta_1 + 2\beta_2 \leq r$ and $\beta_1 \leq 1$. This is precisely the system of polynomial equations (24) if we replace d_j by x_j , a_j by y_j , $2b_j$ by z_j , α_1, α_2 by n_1, n_2 . Now, if we choose $r > \bar{r}(l)$, the system of polynomial equations (53) has only trivial solution $a_j = b_j = 0$ for all $1 \leq j \leq l$. Substitute these results back to system of polynomial equations (21), we also obtain $c_j = 0$ for all $1 \leq j \leq l$, which is a contradiction. This completes our proof.

PROOF OF PROPOSITION 4.3 The proof of part (a) is straightforward from the discussion in Section 4.1. For the proof for part (b), we will present an explicit form for the system of polynomial equations to illustrate the variability of $\underline{\rho}(l)$ and $\bar{\rho}(l)$ based on the values of (m, v) .

(b) As $l = 2$ and $r = 6$, the system of polynomial equations (21) can be rewritten as

$$\begin{aligned}
\sum_{i=1}^3 d_i^2 a_i &= 0, \quad \sum_{i=1}^3 d_i^2 a_i^2 + d_i^2 b_i = 0, \quad \sum_{i=1}^3 -(m^3 + m)d_i^2 a_i^2 + 2vd_i^2 c_i = 0, \\
\sum_{i=1}^3 \frac{1}{3} d_i^2 a_i^3 + d_i^2 a_i b_i &= 0, \quad \sum_{i=1}^3 -(m^3 + m)d_i^2 a_i^3 + 6vd_i^2 a_i c_i = 0, \\
\sum_{i=1}^3 \frac{(m^3 + m)^2}{12v^2} d_i^2 a_i^4 - \frac{m^3 + m}{v} d_i^2 a_i^2 c_i - \frac{m^2 + 1}{vm} d_i^2 b_i c_i + d_i^2 c_i^2 &= 0, \\
\sum_{i=1}^3 \frac{1}{6} d_i^2 a_i^4 + d_i^2 a_i^2 b_i + \frac{1}{2} d_i^2 b_i^2 &= 0, \quad \sum_{i=1}^3 \frac{1}{30} d_i^2 a_i^5 + \frac{1}{3} d_i^2 a_i^3 b_i + \frac{1}{2} d_i^2 a_i b_i^2 = 0, \\
\sum_{i=1}^3 \frac{(m^3 + m)^2}{120v^2} d_i^2 a_i^5 - \frac{(m^3 + m)}{6v} d_i^2 a_i^3 c_i - \frac{m^2 + 1}{2vm} d_i^2 a_i b_i c_i + \frac{1}{2} d_i^2 a_i c_i^2 &= 0, \\
\sum_{i=1}^3 \frac{1}{90} d_i^2 a_i^6 + \frac{1}{12} d_i^2 a_i^4 b_i + \frac{1}{2} d_i^2 a_i^2 b_i^2 + \frac{1}{6} d_i^2 b_i^3 &= 0, \\
\sum_{i=1}^3 \frac{(m^3 + m)^3}{720v^3} d_i^2 a_i^6 + \frac{(m^3 + m)^2}{24v^2} d_i^2 a_i^4 c_i + \frac{m^3 + m}{4v} d_i^2 a_i^2 c_i^2 + \\
\frac{(m^2 + 1)^2}{8v^2 m^2} d_i^2 b_i^2 c_i - \frac{m^2 + 1}{4mv} d_i^2 b_i c_i^2 + \frac{1}{6} d_i^2 c_i^3 &= 0. \tag{54}
\end{aligned}$$

When $r = 4$, the system of polynomial equations (21) contains the first 7 equations in the system of polynomial equations (54). Now, m and v are considered as two additional variables in the above system of polynomial equations. Hence, there are 13 variables with only 7 equations. If we choose $d_1 = d_2 = d_3$ and take the lexicographical ordering $a_1 \succ a_2 \succ a_3 \succ b_1 \succ b_2 \succ b_3 \succ c_1 \succ c_2 \succ c_3 \succ m \succ v$, the Grobner bases (cf. Buchberger [1965]) of the above system of polynomial equations will return a non-trivial solution (due to the complexity of the roots, we will not present them here). As a consequence, $\underline{\rho}(l) \geq 5$ under the case $l = 2$.

For $l = 2$ and $r = 5$, the system of polynomial equations (21) retains the first 9 equations in system (54). It can be checked that if we choose $m = \pm 2, v = 1$, then the system of polynomial equations when $r = 5$ does not have any non-trivial solution (note that, we also use the same lexicographical order as that being used in the case $r = 4$). So, $\underline{\rho}(l) = 5$. However, we can check that the value of $m = \frac{1}{10}$ (close to 0 in general) and $v = 1$ will lead the system of polynomial equations (54) to not having any non-trivial solution. Thus, $\overline{\rho}(l) = 6$. This concludes the proof or part (b) of the proposition.

8.3 Proofs for Section 5

FULL PROOF OF THEOREM 5.2 Here, we shall complete the proof of Theorem 5.2, which is the generalization of the argument in Section 5.1 for a special case for G_0 . Note that, the idea of this generalization is also used to the other settings of $G_0 \notin \mathcal{S}_1$. Now, we consider the possible existence of generic components in G_0 , i.e., there are no homologous sets or symmetry components.

Let $u_1 = 1 < u_2 < \dots < u_{\bar{i}_1} \in [1, k_0 + 1]$ such that $(\frac{v_j^0}{1 + (m_j^0)^2}, \theta_j^0) = (\frac{v_l^0}{1 + (m_l^0)^2}, \theta_l^0)$ and $m_j^0 m_l^0 > 0$ for all $u_i \leq j, l \leq u_{i+1} - 1, 1 \leq i \leq \bar{i}_1 - 1$. The constraint $m_j^0 m_l^0 > 0$ is due to the

conformant property of the homologous sets of G_0 . By definition, we have $|I_{u_i}| = u_{i+1} - u_i$ for all $1 \leq i \leq \bar{i}_1 - 1$ where I_{u_i} denotes the set of all components homologous to component u_i .

To show that G_0 is 1-singular, we construct a sequence of $G \in \mathcal{E}_{k_0}$ such that $(p_i, \theta_i, v_i, m_i) = (p_i^0, \theta_i^0, v_i^0, m_i^0)$ for all $u_2 \leq i \leq k_0$, i.e., all the components of G and G_0 are identical from index u_2 up to k_0 . Hence, in the construction of the components from index u_1 to $u_2 - 1$ of G we consider only the homologous set I_{u_1} of G_0 . Utilizing the argument from the special case proof of Theorem 5.2 in Section 5.1, the construction of the sequence of G is specified by $\Delta\theta_i = \Delta v_i = \Delta p_i = 0$ and $\sum_{i=u_1}^{u_2-1} p_i \Delta m_i / v_i^0 = 0$. Thus G_0 is 1-singular. It remains to demonstrate that $G_0 \in \mathcal{S}_1$ is not 2-singular relative to \mathcal{E}_{k_0} .

Indeed, consider any sequence $G \in \mathcal{E}_{k_0} \rightarrow G_0$ under W_2 distance. Since $W_2^2(G, G_0) \asymp D_2(G_0, G)$ (cf. Lemma 3.1), we have the 2-minimal form for the sequence G as

$$\frac{p_G(x) - p_{G_0}(x)}{W_2^2(G, G_0)} \asymp \frac{A_1(x) + A_2(x)}{D_2(G_0, G)},$$

where $A_1(x)/D_2(G_0, G)$ and $A_2(x)/D_2(G_0, G)$ are linear combinations of the elements of the forms $\frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0)$ for any $1 \leq i \leq k_0$ and $0 \leq |\alpha| \leq 2$. In $A_1(x)/D_2(G_0, G)$, the indices of the components range from 1 to $s_{\bar{i}_1} - 1$. In $A_2(x)/D_2(G_0, G)$, the indices of the components range from $u_{\bar{i}_1}$ to k_0 . It is convenient to think of the term $A_1(x)/D_2(G_0, G)$ as the linear combination of homologous components, and $A_2(x)/D_2(G_0, G)$ as the linear combination of generic components, i.e., no Gaussian nor homologous components.

Regarding $A_2(x)/D_2(G_0, G)$, since we have the system of partial differential equations in (2), the collection of functions in $\left\{ \frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0) : |\alpha| \leq 2, 1 \leq i \leq k_0 \right\}$ are not linearly independent. Employing the same strategy described in Section 4, we obtain a reduced system of linearly independent partial derivatives in Lemma 4.3. This is the set $\left\{ \frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0) : \alpha \in \mathcal{F}_2, 1 \leq i \leq k_0 \right\}$. Let $\lambda_{\alpha_1 \alpha_2 \alpha_3}^{(2)}(\eta_i^0)/D_2(G_0, G)$ be the coefficient of the terms $\frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} v^{\alpha_2} m^{\alpha_3}}(x|\eta_i^0)$ for any $s_{\bar{i}_1} \leq i \leq k_0$ and $\alpha \in \mathcal{F}_2$. The formulae for $\lambda_{\alpha_1, \alpha_2, \alpha_3}^{(2)}$ will be given later in Case 2.

Regarding $A_1(x)/D_2(G_0, G)$, by exploiting the fact that $(\frac{v_j^0}{1 + (m_j^0)^2}, \theta_j^0) = (\frac{v_l^0}{1 + (m_l^0)^2}, \theta_l^0)$ for all $u_i \leq j, l \leq u_{i+1} - 1, 1 \leq i \leq \bar{i}_1 - 1$, the term $A_1(x)/D_2(G_0, G)$ can be written as

$$\begin{aligned} \frac{A_1(x)}{D_2(G_0, G)} = \frac{1}{D_2(G_0, G)} & \left(\sum_{l=1}^{\bar{i}_1-1} \left\{ \sum_{i=u_l}^{u_{l+1}-1} \left[\sum_{j=1}^5 \beta_{jil}^{(2)}(x - \theta_{u_l}^0)^{j-1} \right] f\left(\frac{x - \theta_{u_l}^0}{\sigma_i^0}\right) \Phi\left(\frac{m_i^0(x - \theta_{u_l}^0)}{\sigma_i^0}\right) \right\} + \right. \\ & \left. \left[\sum_{j=1}^4 \gamma_{jil}^{(2)}(x - \theta_{u_l}^0)^{j-1} \right] \exp\left(-\frac{(m_{u_l}^0)^2 + 1}{2v_{u_l}^0}(x - \theta_{u_l}^0)^2\right) \right), \end{aligned}$$

where $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$. (This form is a general version of Eq. (26) in Section (5.1) when $\bar{i}_1 = 2, u_1 = 1, u_2 = k_0 + 1$). The detailed formulas of $\beta_{jil}^{(2)}$ and $\gamma_{jil}^{(2)}$ for $1 \leq l \leq \bar{i}_1 - 1, u_l \leq i \leq u_{l+1} - 1$, and $1 \leq j \leq 5$ are thus similar to that of (26). Here, we rewrite their general fomulations for the transparency of subsequent arguments:

$$\begin{aligned}
\beta_{1il}^{(2)} &= \frac{2\Delta p_i}{\sigma_i^0} - \frac{p_i\Delta v_i}{(\sigma_i^0)^3} - \frac{p_i(\Delta\theta_i)^2}{(\sigma_i^0)^3} + \frac{3p_i(\Delta v_i)^2}{4(\sigma_i^0)^5}, \beta_{2il}^{(2)} = \frac{2p_i\Delta\theta_i}{(\sigma_i^0)^3} - \frac{6p_i\Delta\theta_i\Delta v_i}{(\sigma_i^0)^5}, \\
\beta_{3il}^{(2)} &= \frac{p_i\Delta v_i}{(\sigma_i^0)^5} + \frac{p_i(\Delta\theta_i)^2}{(\sigma_i^0)^5} - \frac{3p_i(\Delta v_i)^2}{2(\sigma_i^0)^7}, \beta_{4il}^{(2)} = \frac{2p_i\Delta\theta_i\Delta v_i}{(\sigma_i^0)^7}, \beta_{5il}^{(2)} = \frac{p_i(\Delta v_i)^2}{4(\sigma_i^0)^9}, \\
\gamma_{1l}^{(2)} &= \sum_{j=u_l}^{u_{l+1}-1} -\frac{p_j m_j^0 \Delta\theta_j}{\pi(\sigma_j^0)^2} + \frac{2p_j m_j^0 \Delta\theta_j \Delta v_j}{\pi(\sigma_j^0)^4} - \frac{2p_j \Delta\theta_j \Delta m_j}{\pi(\sigma_j^0)^2}, \\
\gamma_{2l}^{(2)} &= \sum_{j=s_l}^{s_{l+1}-1} -\frac{p_j m_j^0 \Delta v_j}{2\pi(\sigma_j^0)^4} - \frac{p_j((m_j^0)^3 + 2m_j^0)(\Delta\theta_j)^2}{2\pi(\sigma_j^0)^4} + \frac{p_j \Delta m_j}{\pi(\sigma_j^0)^2} \\
&\quad + \frac{5p_j m_j^0 (\Delta v_j)^2}{8\pi(\sigma_j^0)^6} - \frac{p_j \Delta m_j \Delta v_j}{\pi(\sigma_j^0)^4}, \\
\gamma_{3l}^{(2)} &= \sum_{j=s_l}^{s_{l+1}-1} \frac{p_j(2(m_j^0)^2 + 2)\Delta m_j \Delta\theta_j}{\pi(\sigma_j^0)^4} - \frac{p_j((m_j^0)^3 + 2m_j^0)\Delta\theta_j \Delta v_j}{2\pi(\sigma_j^0)^6}, \\
\gamma_{4l}^{(2)} &= \sum_{j=s_l}^{s_{l+1}-1} -\frac{p_j((m_j^0)^3 + 2m_j^0)(\Delta v_j)^2}{8\pi(\sigma_j^0)^8} - \frac{p_j m_j^0 (\Delta m_j)^2}{2\pi(\sigma_j^0)^4} \\
&\quad + \frac{p_j((m_j^0)^2 + 1)\Delta m_j \Delta v_j}{\pi(\sigma_j^0)^6},
\end{aligned}$$

where $1 \leq l \leq \bar{i}_1 - 1$ and $u_l \leq i \leq u_{l+1} - 1$. Now, suppose that all the coefficients of $A_1(x)/D_2(G_0, G)$ and $A_2(x)/D_2(G_0, G)$ go to 0. It implies that $\gamma_{jl}^{(2)}/D_2(G_0, G)$ ($1 \leq j \leq 4$, $1 \leq l \leq \bar{i}_1 - 1$), $\beta_{jil}^{(2)}/D_2(G_0, G)$ ($1 \leq j \leq 5$, $u_l \leq i \leq u_{l+1} - 1$, $1 \leq l \leq \bar{i}_1 - 1$), and $\lambda_{\alpha_1\alpha_2\alpha_3}^{(2)}(\eta_i^0)/D_2(G_0, G)$ (for all $|\alpha| \leq 2$) go to 0. From the formation of $D_2(G_0, G)$, we can find at least one index $1 \leq i^* \leq k_0$ such that $\left(|\Delta p_{i^*}| + p_{i^*}(|\Delta\theta_{i^*}|^2 + |\Delta v_{i^*}|^2 + |\Delta m_{i^*}|^2)\right)/D_2(G_0, G) \not\rightarrow 0$. Let

$$\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) = |\Delta p_{i^*}| + p_{i^*}(|\Delta\theta_{i^*}|^2 + |\Delta v_{i^*}|^2 + |\Delta m_{i^*}|^2).$$

Now, there are two possible cases for i^* :

Case 1 $u_1 \leq i^* \leq u_{\bar{i}_1} - 1$. Without loss of generality, we assume that $u_1 \leq i^* \leq u_2 - 1$. Denote

$$d(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) = \sum_{j=u_1}^{u_2-1} |\Delta p_j| + p_j(|\Delta\theta_j|^2 + |\Delta v_j|^2 + |\Delta m_j|^2).$$

Since $\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*})/D_2(G_0, G) \not\rightarrow 0$, we have

$$d(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*})/D_2(G_0, G) \not\rightarrow 0.$$

Therefore, for $1 \leq j \leq 5$ and $u_1 \leq i \leq u_2 - 1$, $D_j := \frac{\alpha_{ji1}}{d(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*})} \rightarrow 0$. Now, our argument for this case is organized further into two steps:

Step 1.1 From the vanishes of D_2 and D_4 , we obtain $p_i \Delta \theta_i / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \rightarrow 0$ for all $u_1 \leq i \leq u_2 - 1$. Combining this result with $D_1 \rightarrow 0$ and $D_5 \rightarrow 0$, we achieve for all $u_1 \leq i \leq u_2 - 1$ that

$$\Delta p_i / d(p_i^*, \theta_i^*, v_i^*, m_i^*), p_i \Delta v_i / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \rightarrow 0.$$

Therefore, for all $u_1 \leq i \leq u_2 - 1$, $p_i (\Delta \theta_i)^2 / d(p_i^*, \theta_i^*, v_i^*, m_i^*)$, $p_i (v_i)^2 / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \rightarrow 0$. These results eventually show that

$$U := \left(\sum_{j=u_1}^{u_2-1} p_j (\Delta m_j)^2 \right) / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \not\rightarrow 0.$$

Step 1.2 Since $p_i \Delta \theta_i / d(p_i^*, \theta_i^*, v_i^*, m_i^*)$, $p_i \Delta v_i / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \rightarrow 0$, by using the result that $\gamma_{41}^{(2)} / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \rightarrow 0$, we have

$$V := \left[\sum_{j=u_1}^{u_2-1} \frac{p_j m_j^0 (\Delta m_j)^2}{(\sigma_j^0)^4} \right] / d(p_i^*, \theta_i^*, v_i^*, m_i^*) \rightarrow 0.$$

As $U \not\rightarrow 0$, we obtain

$$V/U = \left[\sum_{j=u_1}^{u_2-1} \frac{p_j m_j^0 (\Delta m_j)^2}{(\sigma_j^0)^4} \right] / \sum_{j=u_1}^{u_2-1} p_j (\Delta m_j)^2 \rightarrow 0. \quad (55)$$

Since $m_i^0 m_j^0 > 0$ for all $u_1 \leq i, j \leq u_2 - 1$, without loss of generality we assume that $m_j^0 > 0$ for all $s_1 \leq j \leq s_2 - 1$. However, it implies that

$$\left[\sum_{j=u_1}^{u_2-1} \frac{p_j m_j^0 (\Delta m_j)^2}{(\sigma_j^0)^4} \right] / \sum_{j=u_1}^{u_2-1} p_j (\Delta m_j)^2 \geq m_{\min} \sum_{j=u_1}^{u_2-1} p_j (\Delta m_j)^2 / \sum_{j=u_1}^{u_2-1} p_j (\Delta m_j)^2, \quad (56)$$

where $m_{\min} := \min_{u_1 \leq j \leq u_2-1} \left\{ \frac{m_j^0}{(\sigma_j^0)^4} \right\}$. Combining with (55), $m_{\min} = 0$ — a contradiction. In sum, Case 1 cannot happen.

Case 2 $u_{i_1}^- \leq i^* \leq k_0$. We can write down the formation of $A_2(x)/D_2(G_0, G)$ as follows

$$\frac{A_2(x)}{D_2(G_0, G)} = \frac{1}{D_2(G_0, G)} \left(\sum_{i=u_{i_1}^-}^{k_0} \sum_{\alpha \in \mathcal{F}_2} \lambda_{\alpha_1, \alpha_2, \alpha_3}^{(2)}(\eta_i^0) \frac{\partial^{|\alpha|} f}{\partial \theta^{\alpha_1} \partial v^{\alpha_2} \partial m^{\alpha_3}}(x | \eta_i^0) \right),$$

where $\lambda_{\alpha_1, \alpha_2, \alpha_3}^{(2)}(\eta_i^0)$ are given by

$$\begin{aligned} \lambda_{0,0,0}^{(2)}(\eta_i^0) &= \Delta p_i, \quad \lambda_{1,0,0}^{(2)}(\eta_i^0) = p_i \Delta \theta_i, \quad \lambda_{0,1,0}^{(2)}(\eta_i^0) = p_i \Delta v_i + p_i (\Delta \theta_i)^2, \\ \lambda_{0,0,1}^{(2)}(\eta_i^0) &= -\frac{(m_1^0)^3 + m_1^0}{2v_1^0} p_i (\Delta \theta_1)^2 - \frac{1}{v_1^0} p_i \Delta v_i \Delta m_i + p_i \Delta m_i, \\ \lambda_{0,2,0}^{(2)}(\eta_i^0) &= p_i (\Delta v_i)^2, \quad \lambda_{0,0,2}^{(2)}(\eta_i^0) = -\frac{(m_1^0)^2 + 1}{2v_1^0 m_1^0} p_i \Delta v_i \Delta m_i + p_i \Delta (m_i)^2, \\ \lambda_{1,1,0}^{(2)}(\eta_i^0) &= p_i \Delta \theta_i \Delta v_i, \quad \lambda_{1,0,1}^{(2)}(\eta_i^0) = p_i \Delta \theta_i \Delta m_i. \end{aligned}$$

From the assumption with the coefficients of $A_2(x)/D_2(G_0, G)$, we have $\lambda_{\alpha_1, \alpha_2, \alpha_3}^{(2)}(\eta_i^0)/D_2(G_0, G) \rightarrow 0$ for any $u_{\bar{i}_1} \leq i \leq k_0$. From the hypothesis with i^* , we have $\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*})/D_2(G_0, G) \not\rightarrow 0$. Therefore, it leads to $\lambda_{\alpha_1, \alpha_2, \alpha_3}^{(2)}(\eta_i^0)/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*})$ for any $u_{\bar{i}_1} \leq i \leq k_0$ and $\alpha \in \mathcal{F}_2$.

Now, since $\lambda_{1,0,0}^{(2)}(\eta_{i^*}^0)/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$, we obtain $\Delta\theta_{i^*}/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$. Combining this result with $\lambda_{1,0,0}^{(2)}(\eta_{i^*}^0)/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$, we have $\Delta v_{i^*}/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$. Furthermore, as $\lambda_{0,0,1}^{(2)}(\eta_{i^*}^0)/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$, we get $\Delta m_{i^*}/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$. Hence, since $\lambda_{0,0,0}^{(2)}(\eta_{i^*}^0)/\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*}) \rightarrow 0$, we ultimately obtain

$$1 = \frac{|\Delta p_{i^*}| + p_{i^*}(|\Delta\theta_{i^*}|^2 + |\Delta v_{i^*}|^2 + |\Delta m_{i^*}|^2)}{\tau(p_{i^*}, \theta_{i^*}, v_{i^*}, m_{i^*})} \rightarrow 0,$$

which is a contradiction. As a consequence, Case 2 cannot happen.

Summarizing, not all the coefficients $\gamma_{jl}^{(2)}/D_2(G_0, G)$ ($1 \leq j \leq 4, 1 \leq l \leq \bar{i}_1 - 1$), $\beta_{jil}^{(2)}/D_2(G_0, G)$ ($1 \leq j \leq 5, u_l \leq i \leq u_{l+1} - 1, 1 \leq l \leq \bar{i}_1 - 1$), $\lambda_{\alpha_1 \alpha_2 \alpha_3}^{(2)}(\eta_i^0)/D_2(G_0, G)$ (for all $\alpha \in \mathcal{F}_2$) go to 0. From Definition 3.2, G_0 is not 2-singular relative to \mathcal{E}_{k_0} . This concludes our proof.

FULL PROOF OF THEOREM 5.3 We divide the proof of this theorem into two main steps.

Step 1: To illustrate our calculations, we consider at first a simple setting of $G_0 \in \mathcal{S}_2$ in which $m_1^0, m_2^0, \dots, m_{k_0}^0 = 0$, leaving out the possible setting of conformant homologous sets and generic components. A complete proof for all possible settings of $G_0 \in \mathcal{S}_2$ will be given in Step 2.

G_0 is 2-singular To establish this, we look at 2-minimal form for $(p_G(x) - p_{G_0}(x))/W_2^2(G, G_0)$, which is asymptotically equal to

$$\frac{1}{W_2^2(G, G_0)} \left[\sum_{i=1}^{k_0} \left(\sum_{j=1}^5 \zeta_{ji}^{(2)}(x - \theta_i^0)^{j-1} \right) f\left(\frac{x - \theta_i^0}{\sigma_i^0}\right) \right], \quad (57)$$

where $\zeta_{li}^{(2)}$ are the polynomials in terms of $\Delta\theta_j, \Delta v_j, \Delta m_j$, and Δp_j as $1 \leq i, j \leq k_0$ and $1 \leq l \leq 5$. To make all the coefficients vanish, it suffices to have $(\Delta v_i)^2/W_2^2(G, G_0) \rightarrow 0$ and

$$\begin{aligned} \left[-\frac{p_i \Delta v_i}{2(\sigma_i^0)^3} - \frac{p_i (\Delta\theta_i)^2}{2(\sigma_i^0)^3} + \frac{3p_i (\Delta v_i)^2}{8(\sigma_i^0)^5} - \frac{2p_i \Delta\theta_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^2} + \frac{\Delta p_i}{\sigma_i^0} \right] / W_2^2(G, G_0) &\rightarrow 0, \\ \left[\frac{\Delta\theta_i}{(\sigma_i^0)^3} + \frac{2\Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^2} - \frac{3\Delta\theta_i \Delta v_i}{2(\sigma_i^0)^5} - \frac{2\Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^4} \right] / W_2^2(G, G_0) &\rightarrow 0, \\ \left[\frac{\Delta v_i}{2(\sigma_i^0)^5} + \frac{(\Delta\theta_i)^2}{2(\sigma_i^0)^5} + \frac{2\Delta\theta_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^4} \right] / W_2^2(G, G_0) &\rightarrow 0, \\ \left[\frac{\Delta\theta_i \Delta v_i}{2(\sigma_i^0)^7} + \frac{\Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^6} \right] / W_2^2(G, G_0) &\rightarrow 0. \end{aligned} \quad (58)$$

This can be achieved by choosing a sequence of $G \rightarrow G_0$ in W_2 such that $\Delta\theta_i = \Delta v_i = \Delta m_i = \Delta p_i = 0$ for all $2 \leq i \leq k_0$; only for component 1 do we set $\Delta\theta_1 = -2\Delta m_1 \sigma_1^0 / \sqrt{2\pi}$ and $\Delta v_1 = (\Delta\theta_1)^2 / 2$. It follows that G_0 is 2-singular relative to \mathcal{E}_{k_0} .

G_0 is not 3-singular The 3-minimal form of $(p_G(x) - p_{G_0}(x))/W_3^3(G, G_0)$ is asymptotically equal to

$$\frac{1}{W_3^3(G, G_0)} \left[\sum_{i=1}^{k_0} \left(\sum_{j=1}^7 \zeta_{ji}^{(3)} (x - \theta_i^0)^{j-1} \right) f \left(\frac{x - \theta_i^0}{\sigma_i^0} \right) \right], \quad (59)$$

where $\zeta_{li}^{(3)}$ are the polynomials in terms of $\Delta\theta_j$, Δv_j , Δm_j , and Δp_j as $1 \leq i, j \leq k_0$ and $1 \leq l \leq 7$. Suppose that there exists a sequence $G \rightarrow G_0$ under W_3 such that all the coefficients of the 3-minimal form vanish. For any $1 \leq i \leq k_0$, it follows after some calculations that

$$\begin{aligned} C_1^{(i)} &:= \left[-\frac{p_i \Delta v_i}{2(\sigma_i^0)^3} - \frac{p_i (\Delta \theta_i)^2}{2(\sigma_i^0)^3} + \frac{3p_i (\Delta v_i)^2}{8(\sigma_i^0)^5} - \frac{2p_i \Delta \theta_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^2} + \frac{3p_i (\Delta \theta_i)^2 \Delta v_i}{4(\sigma_i^0)^5} + \right. \\ &\quad \left. \frac{2p_i \Delta \theta_i \Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^4} + \frac{\Delta p_i}{\sigma_i^0} \right] / W_3^3(G, G_0) \rightarrow 0, \\ C_2^{(i)} &:= \left[\frac{p_i \Delta \theta_i}{(\sigma_i^0)^3} + \frac{2p_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^2} - \frac{3p_i \Delta \theta_i \Delta v_i}{2(\sigma_i^0)^5} - \frac{2p_i \Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^4} - \frac{p_i (\Delta \theta_i)^3}{2(\sigma_i^0)^5} - \right. \\ &\quad \left. \frac{3p_i (\Delta \theta_i)^2 \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^4} + \frac{15p_i \Delta \theta_i (\Delta v_i)^2}{8(\sigma_i^0)^7} + \frac{2p_i (\Delta v_i)^2 \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^6} \right] / W_3^3(G, G_0) \rightarrow 0, \\ C_3^{(i)} &:= \left[\frac{p_i \Delta v_i}{2(\sigma_i^0)^5} + \frac{p_i (\Delta \theta_i)^2}{2(\sigma_i^0)^5} - \frac{3p_i (\Delta v_i)^2}{4(\sigma_i^0)^7} + \frac{2p_i \Delta \theta_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^4} - \frac{3p_i (\Delta \theta_i)^2 \Delta v_i}{2(\sigma_i^0)^7} - \right. \\ &\quad \left. \frac{5\Delta \theta_i \Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^6} \right] / W_3^3(G, G_0) \rightarrow 0, \\ C_4^{(i)} &:= \left[\frac{p_i \Delta \theta_i \Delta v_i}{2(\sigma_i^0)^7} + \frac{p_i \Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^6} + \frac{p_i (\Delta \theta_i)^3}{6(\sigma_i^0)^7} - \frac{p_i (\Delta m_i)^3}{3\sqrt{2\pi}(\sigma_i^0)^4} + \right. \\ &\quad \left. \frac{p_i (\Delta \theta_i)^2 \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^6} - \frac{5p_i \Delta \theta_i (\Delta v_i)^2}{4(\sigma_i^0)^9} - \frac{2p_i (\Delta v_i)^2 \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^8} \right] / W_3^3(G, G_0) \rightarrow 0, \\ C_5^{(i)} &:= \left[\frac{p_i (\Delta v_i)^2}{8(\sigma_i^0)^9} - \frac{5p_i (\Delta v_i)^3}{16(\sigma_i^0)^{11}} + \frac{p_i (\Delta \theta_i)^2 \Delta v_i}{4(\sigma_i^0)^9} + \frac{p_i \Delta \theta_i \Delta v_i \Delta m_i}{\sqrt{2\pi}(\sigma_i^0)^8} \right] / W_3^3(G, G_0) \rightarrow 0, \\ C_6^{(i)} &:= \left[\frac{p_i \Delta \theta_i (\Delta v_i)^2}{8(\sigma_i^0)^{11}} + \frac{p_i (\Delta v_i)^2 \Delta m_i}{4\sqrt{2\pi}(\sigma_i^0)^{10}} \right] / W_3^3(G, G_0) \rightarrow 0, \\ C_7^{(i)} &:= p_i (\Delta v_i)^3 / 48(\sigma_i^0)^3 W_3^3(G, G_0) \rightarrow 0. \quad (60) \end{aligned}$$

Since the system of limits in (60) holds for any $1 \leq i \leq k_0$, to further simplify the argument without loss of generality, we consider $k_0 = 1$. Under that scenario, we can rewrite $W_3^3(G, G_0) = p_1(|\Delta\theta_1|^3 + |\Delta v_1|^3 + |\Delta m_1|^3)$ where $p_1 = 1$. Additionally, for the simplicity of the presentation, we denote $C_i := C_i^{(1)}$ for any $1 \leq i \leq 7$. Now, our argument is organized into the following key steps

Step 1.1: We will argue that $\Delta\theta_1, \Delta v_1, \Delta m_1 \neq 0$. If $\Delta\theta_1 = 0$, by combining the vanishing of C_5 and C_7 , we achieve $(\Delta v_1)^2 / W_3^3(G, G_0) \rightarrow 0$. Combining this result with $C_3 \rightarrow 0$, we obtain $\Delta v_1 / W_3^3(G, G_0) \rightarrow 0$. Combining the previous results with $C_4 \rightarrow 0$ eventually yields that $(\Delta m_1)^3 / W_3^3(G, G_0) \rightarrow 0$. Hence, $1 = p_1(|\Delta v_1|^3 + |\Delta m_1|^3) / W_3^3(G, G_0) \rightarrow 0$, which is a contradiction.

If $\Delta v_1 = 0$, then $C_1 + \Delta\theta_1 C_2 \rightarrow 0$ implies that $(\Delta\theta_1)^2 / W_3^3(G, G_0) \rightarrow 0$. Combining this result with $C_4 \rightarrow 0$, we achieve $(\Delta m_1)^3 / W_3^3(G, G_0) \rightarrow 0$, which also leads to a contradiction.

If $\Delta m_1 = 0$, then $C_6 \rightarrow 0$ leads to $(\Delta\theta_1)(\Delta v_1)^2/W_3^3(G, G_0) \rightarrow 0$. Combine this result with $C_4 \rightarrow 0$ leads to

$$\left[\frac{(\Delta\theta_1)(\Delta v_1)}{2(\sigma_1^0)^7} + \frac{(\Delta\theta_1)^3}{6(\sigma_1^0)^7} \right] / W_3^3(G, G_0) \rightarrow 0. \quad (61)$$

The combination of the above result and $C_3 \rightarrow 0$ implies that $\Delta v_1/W_3^3(G, G_0) \rightarrow 0$. Combine the former result with (61), we obtain $(\Delta\theta_1)^3/W_3^3(G, G_0) \rightarrow 0$, which is also a contradiction. Overall, we obtain the conclusion of this step.

Step 1.2: If $|\Delta v_1|$ is the maximum among $|\Delta\theta_1|$, $|\Delta v_1|$, and $|\Delta m_1|$. Then from $C_7 \rightarrow 0$, we obtain $|\Delta v_1|^3/(|\Delta\theta_1|^3 + |\Delta v_1|^3 + |\Delta m_1|^3) \rightarrow 0$, which is a contradiction.

Step 1.3: If $|\Delta\theta_1|$ is the maximum among $|\Delta\theta_1|$, $|\Delta v_1|$, and $|\Delta m_1|$. Denote $\Delta v_1/\Delta\theta_1 \rightarrow k_1$ and $\Delta m_1/\Delta\theta_1 \rightarrow k_2$. From C_7 , we obtain $k_1 = 0$. As $C_2 \rightarrow 0$, we obtain

$$\left[-\Delta\theta_1/(\sigma_1^0)^3 + 2\Delta m_1/\sqrt{2\pi}(\sigma_1^0)^2 \right] / (|\Delta\theta_1| + |\Delta v_1| + |\Delta m_1|) \rightarrow 0.$$

By diving both the numerator and denominator of this ratio by $\Delta\theta_1$, we quickly obtain the equation $1/(\sigma_1^0)^3 + 2k_2/\sqrt{2\pi}(\sigma_1^0)^2 = 0$, which yields the solution $k_2 = -\sqrt{\pi}/\sqrt{2}\sigma_1^0$.

Now, $C_5 \rightarrow 0$ yields that $(\Delta v_1)^2/(|\Delta\theta_1|^3 + |\Delta v_1|^3 + |\Delta m_1|^3) \rightarrow 0$. Applying this result to $C_3 \rightarrow 0$ and $C_4 \rightarrow 0$, we have $M_1, M_2 \rightarrow 0$ where the formations of M_1, M_2 are as follows:

$$\begin{aligned} M_1 &:= \left(\frac{\Delta v_1}{2(\sigma_1^0)^5} + \frac{(\Delta\theta_1)^2}{2(\sigma_1^0)^5} + \frac{2(\Delta\theta_1)(\Delta m_1)}{\sqrt{2\pi}(\sigma_1^0)^4} \right) / (|\Delta\theta_1|^3 + |\Delta v_1|^3 + |\Delta m_1|^3), \\ M_2 &:= \left(\frac{(\Delta\theta_1)(\Delta v_1)}{2(\sigma_1^0)^7} + \frac{(\Delta v_1)(\Delta m_1)}{\sqrt{2\pi}(\sigma_1^0)^6} + \frac{(\Delta\theta_1)^3}{6(\sigma_1^0)^7} - \frac{(\Delta m_1)^3}{3\sqrt{2\pi}(\sigma_1^0)^4} + \right. \\ &\quad \left. + \frac{(\Delta\theta_1)^2(\Delta m_1)}{\sqrt{2\pi}(\sigma_1^0)^6} \right) / (|\Delta\theta_1|^3 + |\Delta v_1|^3 + |\Delta m_1|^3). \end{aligned}$$

Now, $\left(\frac{\Delta\theta_1}{(\sigma_1^0)^2} + \frac{2\Delta m_1}{\sqrt{2\pi}\sigma_1^0} \right) M_1 - M_2$ yields that

$$\left[\frac{(\Delta m_1)^3}{3\sqrt{2\pi}} + \frac{2(\theta_1)(\Delta m_1)^2}{\pi\sigma_1^0} + \frac{2(\Delta\theta_1)^2(\Delta m_1)}{\sqrt{2\pi}(\sigma_1^0)^2} + \frac{(\Delta\theta_1)^3}{3(\sigma_1^0)^3} \right] / (|\Delta\theta_1|^3 + |\Delta v_1|^3 + |\Delta m_1|^3) \rightarrow 0.$$

By dividing both the numerator and denominator of this term by $(\Delta\theta_1)^3$, we obtain the equation $\frac{k_2^3}{3\sqrt{2\pi}} + \frac{2k_2^2}{\pi\sigma_1^0} + \frac{2k_2}{\sqrt{2\pi}(\sigma_1^0)^2} + \frac{1}{3(\sigma_1^0)^3} = 0$. Since $k_2 = -\frac{\sqrt{\pi}}{\sqrt{2}\sigma_1^0}$, this equation yields $\pi/6 - 1/3 = 0$, which is a contradiction. Therefore, this step cannot hold.

Step 1.4: If $|\Delta m_1|$ is the maximum among $|\Delta\theta_1|$, $|\Delta v_1|$, and $|\Delta m_1|$. The argument in this step is similar to that of Step 1.3. In fact, by denoting $\Delta\theta_1/\Delta m_1 \rightarrow k_3$ and $\Delta v_1/\Delta m_1 \rightarrow k_4$ then we also achieve $k_4 = 0$ and $k_3 = -\frac{\sqrt{2}}{\sqrt{\pi}\sigma_1^0}$ (by $C_2 \rightarrow 0$). Now by using the limits $C_3, C_4 \rightarrow 0$ as that of Step

1.3 and after some calculations, we obtain the equation $\frac{k_3^3}{3(\sigma_1^0)^3} + \frac{2k_3^2}{\sqrt{2\pi}(\sigma_1^0)^2} + \frac{2k_3}{\pi\sigma_1^0} + \frac{1}{3\sqrt{2\pi}} = 0$, which also does not admit $k_3 = -\frac{\sqrt{2}}{\sqrt{\pi}\sigma_1^0}$ as a solution — a contradiction.

In sum, we have shown under that simple setting of $G_0 \in \mathcal{S}_2$, it is 2-singular, but not 3-singular relative to \mathcal{E}_{k_0} . Therefore, $\ell(G_0|\mathcal{E}_{k_0}) = 2$.

Step 2: Now, we address the general setting of $G_0 \in \mathcal{S}_2$, which accounts for the possible presence of both generic components and conformant homologous sets. Without loss of generality, we assume that $m_1^0, m_2^0, \dots, m_{\bar{i}_2}^0 = 0$ where $1 \leq \bar{i}_2 \leq k_0$ denotes the largest index i such that $m_i^0 = 0$. The remaining components are either conformant homologous sets or generic components. Using the exact same construction as that of Step 1, we establish easily that G_0 is 2-singular relative to \mathcal{E}_{k_0} . It remains to show that G_0 is not 3-singular relative to \mathcal{E}_{k_0} .

Consider the 3-minimal form for any sequence $G \in \mathcal{E}_{k_0} \rightarrow G_0$ under W_3 distance. Since $W_3^3(G, G_0) \asymp D_3(G_0, G)$ (cf. Lemma 3.1), we have

$$\frac{p_G(x) - p_{G_0}(x)}{W_3^3(G, G_0)} \asymp \frac{A'_1(x) + A'_2(x)}{D_3(G_0, G)},$$

where $A'_1(x)/D_3(G_0, G)$ is the linear combination of Gaussian components, i.e., the indices of components range from 1 to \bar{i}_2 , while $A'_2(x)/D_3(G_0, G)$ is the linear combination of conformant homologous components and generic components.

Suppose that all the coefficients of $A'_1(x)/D_3(G_0, G)$, $A'_2(x)/D_3(G_0, G)$ go to 0. Similar to the argument in the proof of Theorem 5.2, observe that there is some index $\underline{i} \in [1, k_0]$ such that $(|\Delta p_{\underline{i}}| + p_{\underline{i}}(|\Delta \theta_{\underline{i}}|^3 + |\Delta v_{\underline{i}}|^3 + |\Delta m_{\underline{i}}|^3))/D_3(G_0, G) \not\rightarrow 0$. There are two possible cases regarding \underline{i} .

Case 2.1 $\underline{i} \in [1, \bar{i}_2]$. Applying a similar argument as that from Step 1 of this proof where we have only Gaussian components, we conclude that not all of the coefficients of $A'_1(x)/D_3(G_0, G)$ vanish, which is a contradiction. Therefore, Case 2.1 cannot happen.

Case 2.2 $\underline{i} \in [\bar{i}_2 + 1, k_0]$. Define

$$D_{r, \text{new}}(G_0, G) = \sum_{i=\bar{i}_2+1}^{k_0} (|\Delta p_i| + p_i(|\Delta \theta_i|^r + |\Delta v_i|^r + |\Delta m_i|^r)),$$

for any $r \in \{2, 3\}$. The idea of $D_{r, \text{new}}(G_0, G)$ is that we truncate the value of $D_r(G_0, G)$ from the index 1 to \bar{i}_2 , i.e., all the indices correspond to Gaussian components.

It is clear that $D_{3, \text{new}}(G_0, G) \lesssim D_{2, \text{new}}(G_0, G)$. Since $D_{3, \text{new}}(G_0, G)/D_3(G_0, G) \not\rightarrow 0$, we have $D_{2, \text{new}}(G_0, G)/D_3(G_0, G) \not\rightarrow 0$. By multiplying all the coefficients of $A'_2(x)/D_3(G_0, G)$ with $D_{2, \text{new}}(G_0, G)/D_3(G_0, G)$, we eventually obtain all the coefficients of $A'_2(x)/D_{2, \text{new}}(G_0, G)$ go to 0. However, by utilizing the same argument as in the proof of Theorem 5.2, we reach to the conclusion that the second order Taylor expansion is sufficient to have all the coefficients of $A'_2(x)/D_{2, \text{new}}(G_0, G)$ not vanish. Thus, not all the coefficients of $A'_2(x)/D_3(G_0, G)$ go to 0, which is a contradiction. As a consequence, Case 2.2 also cannot happen.

In sum, under no circumstance can all the coefficients of $A'_1(x)/D_3(G_0, G)$ and $A'_2(x)/D_3(G_0, G)$ be made to vanish. Hence, $G_0 \in \mathcal{S}_2$ is not 3-singular relative to \mathcal{E}_{k_0} , which concludes the proof.

8.4 Proofs for Section 7

PROOF OF PROPOSITION 7.1 (a) The proof proceeds by induction on l . When $l = 1$, the conclusion clearly holds. Assume that that conclusion of the proposition holds for $l - 1$. We will demonstrate that it also holds for l . Denote $y_i = a_i c_i$ and $z_i = b_i c_i$ for all $1 \leq i \leq l + 1$. Then, we can

rewrite system of polynomial equations (36) as follows: $\sum_{i=1}^{l+1} z_i^u y_i = 0$ for any $0 \leq u \leq l$. If there exists some $1 \leq i_1 \leq l+1$ such that $c_{i_1} = 0$, then we go back to the case $l-1$, which we have already known from the hypothesis that we do not have non-trivial solution. Therefore, we assume that $c_i \neq 0$ for all $1 \leq i \leq l+1$, which implies that $y_i \neq 0$ for all $1 \leq i \leq l+1$. Now, the system of equations has the

form of Vardermode matrix, which is $\begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^s & z_2^s & \dots & z_{l+1}^s \end{bmatrix}$. By suitable linear transformations, we

can rewrite the original system of equations as the following equivalent equations $\prod_{j \neq i} (z_j - z_i) y_i = 0$ for all $1 \leq i \leq l+1$. Since $y_i \neq 0$ for all $1 \leq i \leq l+1$, we obtain $\prod_{j \neq i} (z_j - z_i) = 0$ for all $1 \leq i \leq l+1$. As a consequence, there exists a partition J_1, J_2, \dots, J_s of $\{1, 2, \dots, l+1\}$ for some $1 \leq s \leq [l/2]$ such that if $i_2, i_3 \in J_u$ for $1 \leq u \leq s$, we have $z_{i_2} = z_{i_3}$ and for any $1 \leq i \neq j \leq s$, any two elements $z_{i_4} \in J_i, z_{j_4} \in J_j$ are different. Choose any $j_i \in J_i$ for all $1 \leq i \leq s$. It is clear that the system of equations can be rewritten as $\sum_{i=1}^s z_{j_i}^u \sum_{j \in J_i} y_j = 0$ for all $0 \leq u \leq l+1$. If $s \geq 2$, it indicates that $|J_i| \leq l$ for all $1 \leq i \leq s$. Now, if we have some $1 \leq i_4 \leq s$ such that $\sum_{j \in J_{i_4}} y_j = 0$ then we obtain $\sum_{j \in J_{i_4}} a_j c_j = 0$. Since $z_{i_1} = z_{i_2}$ for any $i_1, i_2 \in J_{i_4}$, this equation can be rewritten as $\sum_{j \in J_{i_4}} a_j \prod_{v \neq j} b_v = 0$, which is a contradiction to the assumption of part (a) of the proposition. Therefore, $\sum_{j \in J_i} y_j \neq 0$ for all $1 \leq i \leq s$. However, by using the same argument as before, again by linear transformation, we can rewrite the new system of polynomial equations as $\sum_{j \in J_i} y_j \prod_{v \neq i} (z_{j_u} - z_{j_i}) = 0$ for all $1 \leq i \leq s$. This implies that there should be some $1 \leq u_1 \neq u_2 \leq s$ such that $z_{j_{u_1}} = z_{j_{u_2}}$, which is a contradiction.

As a consequence, we have $s = 1$, i.e., $|I_1| = l+1$. Hence, $b_1 c_1 = b_2 c_2 = \dots = b_{l+1} c_{l+1}$. Combining this fact with the equation $\sum_{i=1}^{l+1} a_i c_i = 0$, we obtain $\sum_{i=1}^{l+1} a_i \prod_{j \neq i} b_j = 0$, which is a contradiction to the assumption of the proposition. This concludes the proof.

(b) We choose $c_i = 0$ for all $i \notin I \subset \{1, \dots, l\}$. The system of polynomial equations (36) becomes $\sum_{i \in I} a_i b_i^u c_i^{u+1} = 0$ for all $u \geq 0$. Notice that by choosing $b_i c_i = b_j c_j$ for all $i, j \in I$, we have $\sum_{i \in I} a_i b_i^u c_i^{u+1} = b_j c_j \sum_{i \in I} a_i c_i = 0$ for some $j \in I$ and for all $u \geq 1$ as long as $\sum_{i \in I} a_i c_i = 0$. Combining all the conditions, we obtain $\sum_{i \in I} a_i \prod_{j \neq i} b_j = 0$, which completes the proof.

(c) The result for the case $l = 1$ is obvious. For the case $l = 2$, after replacing c_3 in terms of c_1, c_2 , we obtain the following quadratic equation $(a_1 a_3 b_1 + a_1^2 b_3) c_1^2 + 2a_1 a_2 b_3 c_1 c_2 + (a_2 a_3 b_2 + a_2^2 b_3) c_2^2 = 0$. Note that, $c_1, c_2 \neq 0$ due to the assumption of part (c). Therefore, we does not have solution of this quadratic equation when $a_1^2 a_2^2 b_3^2 < (a_1 a_3 b_1 + a_1^2 b_3)(a_2 a_3 b_2 + a_2^2 b_3)$. It is equivalent to $\sum_{i=1}^3 a_i \prod_{j \neq i} b_j > 0$, which confirms our hypothesis. We are done.

FULL PROOF OF THEOREM 7.1 Here, we only provide the proof for part (b) as the proof for part (a) is similar. This is a generalization of the argument in Section 7.1. Under this situation, apart from the nonconformant homologous sets without $C(1)$ singularity, we also have for G_0 the presence

of Gaussian components components and possibly some conformant homologous sets, in addition to some generic components.

Let $u_1 = 1 < u_2 < \dots < u_{\bar{i}_3} \in [1, k_0 + 1]$ such that $(\frac{v_j^0}{1 + (m_j^0)^2}, \theta_j^0) = (\frac{v_l^0}{1 + (m_l^0)^2}, \theta_l^0)$ for all $u_i \leq j, l \leq u_{i+1} - 1, 1 \leq i \leq \bar{i}_3 - 1$, i.e., all the nonconformant homologous components without type C(1) singularity are from index 1 to $u_{\bar{i}_3}$. The remaining components are either Gaussian ones or conformant homologous sets or generic ones. It follows that $|I_{u_i}| = u_{i+1} - u_i$ for all $1 \leq i \leq \bar{i}_3 - 1$ and all I_{u_i} are nonconformant homologous sets without C(1) singularity.

Consider the \bar{r} -th minimal form for any sequence $G \in \mathcal{E}_{k_0} \rightarrow G_0$ under $W_{\bar{r}}$ distance where $\bar{r} = \max \left\{ 3, \bar{s}(G_0) + 1 \right\}$. Since $W_{\bar{r}}^{\bar{r}}(G, G_0) \asymp D_{\bar{r}}(G_0, G)$ (cf. Lemma 3.1), we have

$$\frac{p_G(x) - p_{G_0}(x)}{W_{\bar{r}}^{\bar{r}}(G, G_0)} \asymp \frac{B_1(x) + B_2(x)}{D_{\bar{r}}(G_0, G)},$$

where $B_1(x)/D_{\bar{r}}(G_0, G)$ is the linear combination of nonconformant homologous components, i.e., the indices of components range from 1 to \bar{i}_3 while $B_2(x)/D_{\bar{r}}(G_0, G)$ is the linear combination of conformant homologous components, Gaussian components, and generic components.

Now, suppose that all the coefficients of $B_1(x)/D_{\bar{r}}(G_0, G), B_2(x)/D_{\bar{r}}(G_0, G)$ go to 0. Similar to the argument employed in the proof of Theorem 5.2, there is some index $\underline{i} \in [1, k_0]$ such that $(|\Delta p_{\underline{i}}| + p_{\underline{i}}(|\Delta \theta_{\underline{i}}|^{\bar{r}} + |\Delta v_{\underline{i}}|^{\bar{r}} + |\Delta m_{\underline{i}}|^{\bar{r}}))/D_{\bar{r}}(G_0, G) \not\rightarrow 0$. Now, there are two possible scenarios regarding \underline{i}

Case 1.1 $\underline{i} \in [1, u_{\bar{i}_3} - 1]$. Under that case, we can check that

$$\begin{aligned} \frac{B_1(x)}{D_{\bar{r}}(G_0, G)} = \frac{1}{D_{\bar{r}}(G_0, G)} & \left(\sum_{l=1}^{\bar{i}_3-1} \left\{ \sum_{i=u_l}^{u_{l+1}-1} \left[\sum_{j=1}^{2\bar{r}+1} \beta_{jil}^{(\bar{r})} (x - \theta_{u_l}^0)^{j-1} \right] f\left(\frac{x - \theta_{u_l}^0}{\sigma_i^0}\right) \Phi\left(\frac{m_i^0(x - \theta_{u_l}^0)}{\sigma_i^0}\right) \right\} + \right. \\ & \left. \left[\sum_{j=1}^{2\bar{r}} \gamma_{jl}^{(\bar{r})} (x - \theta_{u_l}^0)^{j-1} \right] \exp\left(-\frac{(m_{u_l}^0)^2 + 1}{2v_{u_l}^0} (x - \theta_{u_l}^0)^2\right) \right). \end{aligned}$$

This representation of $B_1(x)/D_{\bar{r}}(G_0, G)$ is the general formulation of the equation (26) in Section (5.1) where $\bar{i}_3 = 2, u_1 = 1, u_2 = k_0 + 1$, and $\bar{r} = r$. Since $\underline{i} \in [1, u_{\bar{i}_3} - 1]$, there exists some index $l^* \in [1, \bar{i}_3 - 1]$ such that $\underline{i} \in [u_{l^*}, u_{l^*+1} - 1]$. By means of the same argument as that of Section 7.1 for $\beta_{jil}^{(\bar{r})}/D_{\bar{r}}(G_0, G) \rightarrow 0$ and $\gamma_{jl}^{(\bar{r})}/D_{\bar{r}}(G_0, G) \rightarrow 0$, we can extract the following system of polynomial limits:

$$\sum_{i=u_{l^*}}^{u_{l^*+1}-1} p_i^0(m_i^0)^{l/2-1} (k_i)^{l/2} = 0,$$

where at least one of k_i differs from 0. Here, l is any even number such that $2 \leq l \leq 2\bar{r}$. From the formulation of $\bar{s}(G_0)$, since $\bar{r} \geq \bar{s}(G_0) + 1 \geq \bar{s}(|I_{u_{l^*}}|, \{p_i^0\}_{i \in I_{u_{l^*}}}, \{m_i^0\}_{i \in I_{u_{l^*}}}) + 1$, we can guarantee that the above system of polynomial equations does not have any non-trivial solution, which is a contradiction. Therefore, Case 1.1 cannot happen.

Case 1.2 $i \in [u_{\bar{\tau}_3}, k_0]$. Using the same argument as that in the proof of Theorem 5.3, the third order Taylor expansion is sufficient so that not all the coefficients of $B_2(x)/D_{3,new}(G_0, G)$ go to 0 where

$$D_{3,new}(G_0, G) = \sum_{i=u_{\bar{\tau}_3}}^{k_0} (|\Delta p_i| + p_i(|\Delta \theta_i|^3 + |\Delta v_i|^3 + |\Delta m_i|^3)).$$

Since $\bar{\tau} \geq 3$, we have $D_{3,new}(G_0, G)/D_{\bar{\tau}}(G_0, G) \not\rightarrow 0$. As all the coefficients of $B_2(x)/D_{\bar{\tau}}(G_0, G)$ vanish, it leads to all the coefficients of $B_2(x)/D_{3,new}(G_0, G)$ go to 0, which is a contradiction. Thus, Case 1.2 cannot happen.

In sum, for any sequence of G tending to G_0 in $W_{\bar{\tau}}$, not all the coefficients of $B_1(x)/D_{\bar{\tau}}(G_0, G)$ and $B_2(x)/D_{\bar{\tau}}(G_0, G)$ go to 0. By Definition 3.2, we conclude that $G_0 \in \mathcal{S}_2$ is not $\bar{\tau}$ -singular relative to \mathcal{E}_{k_0} . As a consequence, $\ell(G_0|\mathcal{E}_{k_0}) \leq \bar{\tau} - 1 = \max \left\{ 2, \bar{s}(G_0) \right\}$.

PROOF OF PROPOSITION 7.2 Here, we utilize the same assumption on G_0 as that in the proof of Theorem 7.1, i.e., all the nonconformant homologous sets without C(1) singularity are from index 1 to $u_{\bar{\tau}_3}$. We also rearrange the components of G_0 such that the first nonconformant homologous set without C(1) singularity I_{u_1} has exactly k^* elements, i.e., $u_2 - u_1 = k^*$. As $u_1 = 1$, we have $u_2 = k^* + 1$.

(a) We will demonstrate that G_0 is 1-singular relative to \mathcal{E}_{k_0} . Indeed, the sequence of G is constructed as follows: $p_i = p_i^0, \theta_i = \theta_i^0, v_i = v_i^0$ for all $u_2 = k^* + 1 \leq i \leq k_0$, i.e., we match all the components of G and G_0 except the first k^* components of G_0 . Now, by proceeding in the same way as described in Section 7.1 up to Eq. (33), to verify that G_0 is indeed 1-singular, the choice of the first k^* components of G needs to satisfy

$$\sum_{i=u_1}^{u_2-1} q_i \Delta t_i / \sum_{i=u_1}^{u_2-1} q_i |\Delta t_i| \rightarrow 0,$$

where $q_i = p_i/\sigma_i^0$ and $\Delta t_i = \Delta m_i/\sigma_i^0$ as $u_1 \leq i \leq u_2 - 1$. A simple choice is to take the first k^* components of G by $\sum_{i=u_1}^{u_2-1} q_i \Delta t_i = q_1 \Delta t_1 + q_2 \Delta t_2 = 0$, which is always possible. We conclude that G_0 is 1-singular relative to \mathcal{E}_{k_0} . Since $\bar{s}(G_0) = 1$ as $k^* = 2$, by combining with the upper bound of Theorem 7.1, we have $\ell(G_0|\mathcal{E}_{k_0}) = 1$.

(b) There are two cases to consider in this part

Case 1: All the homologous sets I of G_0 such that $|I| = k^*$ satisfy $\sum_{i \in I} p_i^0 \prod_{j \in I \setminus \{i\}} m_j^0 > 0$. To demonstrate that G_0 is 1-singular relative to \mathcal{E}_{k_0} , we utilize the same construction of G as that of part (a), i.e., $p_i = p_i^0, \theta_i = \theta_i^0, v_i = v_i^0$ for all $u_2 = k^* + 1 \leq i \leq k_0$ and $\sum_{i=u_1}^{u_2-1} q_i \Delta t_i = 0$. Next, we will show that G_0 is not 2-singular relative to \mathcal{E}_{k_0} . Using the same argument as that of the proof of Theorem 7.1, we obtain the following system of limiting rational polynomial functions:

$$\begin{aligned} \sum_{i=u_{l^*}}^{u_{l^*+1}-1} q_i \Delta t_i / \sum_{i=u_{l^*}}^{u_{l^*+1}-1} q_i |\Delta t_i|^2 &\rightarrow 0, \\ \sum_{i=u_{l^*}}^{u_{l^*+1}-1} q_i t_i^0 (\Delta t_i)^2 / \sum_{i=u_{l^*}}^{u_{l^*+1}-1} q_i |\Delta t_i|^2 &\rightarrow 0, \end{aligned}$$

where l^* is some index in $[1, \bar{i}_3 - 1]$ and $q_i = p_i/\sigma_i^0$, $\Delta t_i = \Delta m_i/\sigma_i^0$, $t_i^0 = m_i^0/\sigma_i^0$ for all $u_{l^*} \leq i \leq u_{l^*+1} - 1$. By employing the greedy extraction technique being described in Section 7.1.1, we obtain the following system of polynomial equations:

$$\sum_{i=u_{l^*}}^{u_{l^*+1}-1} p_i^0 c_i = 0, \quad \sum_{i=u_{l^*}}^{u_{l^*+1}-1} p_i^0 m_i^0 c_i^2 = 0,$$

where at least one of c_i differs from 0. Now, we have two possible scenarios:

Case 1.1: $|I_{u_{l^*}}| = u_{l^*+1} - u_{l^*} = 2$. Then, by solving the above system of equations, we obtain $\sum_{i \in I_{u_{l^*}}} p_i^0 \prod_{j \in I_{u_{l^*}} \setminus \{i\}} m_j^0 = 0$, which means $I_{u_{l^*}}$ is nonconformant homologous set with $C(1)$ singularity of G_0 — a contradiction to the fact that $G_0 \in \mathcal{S}_{31}$.

Case 1.2: $|I_{u_{l^*}}| = u_{l^*+1} - u_{l^*} = k^* = 3$. Then, by solving the above system of equations, we obtain $\sum_{i \in I_{u_{l^*}}} p_i^0 \prod_{j \in I_{u_{l^*}} \setminus \{i\}} m_j^0 < 0$ — a contradiction to the assumption of Case 1.

Thus, G_0 is not 2-singular relative to \mathcal{E}_{k_0} . As a consequence, $\ell(G_0|\mathcal{E}_{k_0}) = 1$ under Case 1.

Case 2: There exists at least one nonconformant homologous set I of G_0 such that $|I| = k^*$ satisfies $\sum_{i \in I} p_i^0 \prod_{j \in I \setminus \{i\}} m_j^0 < 0$. Without loss of generality, we assume the homologous set I_{u_1} of G_0 to have the property $\sum_{i \in I_{u_1}} p_i^0 \prod_{j \in I_{u_1} \setminus \{i\}} m_j^0 < 0$. We will show that G_0 is 2-singular relative to \mathcal{E}_{k_0} . In fact, we construct the sequence of G by letting $p_i = p_i^0$, $\theta_i = \theta_i^0$, $v_i = v_i^0$ for all $u_2 = k^* + 1 \leq i \leq k_0$. In order for G_0 to be 2-singular, it is sufficient that

$$\begin{aligned} \sum_{i=u_1}^{u_2-1} q_i \Delta t_i / \sum_{i=u_1}^{u_2-1} q_i |\Delta t_i|^2 &\rightarrow 0, \\ \sum_{i=u_1}^{u_2-1} q_i t_i^0 (\Delta t_i)^2 / \sum_{i=u_1}^{u_2-1} q_i |\Delta t_i|^2 &\rightarrow 0. \end{aligned}$$

The simple solution to the above system of limits is $\sum_{i=u_1}^{u_2-1} q_i \Delta t_i = 0$ and $\sum_{i=u_1}^{u_2-1} q_i t_i^0 (\Delta t_i)^2 = 0$. One solution to these two equations is $p_i = p_i^0$ and $\Delta m_i = (\sigma_i^0)^2 d_i/n$ for all $u_1 \leq i \leq u_2 - 1$ where d_1, d_2, d_3 satisfy

$$\sum_{i=u_1}^{u_2-1} p_i^0 d_i = 0, \quad \sum_{i=u_1}^{u_2-1} p_i^0 m_i^0 d_i^2 = 0,$$

which is guaranteed to have non-trivial solution as $\sum_{i \in I_{u_1}} p_i^0 \prod_{j \in I_{u_1} \setminus \{i\}} m_j^0 < 0$. Therefore, G_0 is 2-singular relative to \mathcal{E}_{k_0} . Since $\overline{s}(G_0) = 2$ as $k^* = 3$, combining with the upper bound of Theorem 7.1, we obtain $\ell(G_0|\mathcal{E}_{k_0}) = 2$ under Case 2. This concludes our proof.

FULL PROOF OF THEOREM 7.2 Here, we shall provide the complete proof of Theorem 7.2, which is also the generalization of the argument in Section 7.1.3. Indeed, without loss of generality, we assume that $(p_1^0/\sigma_1^0, m_1^0/\sigma_1^0) = (p_2^0/\sigma_2^0, -m_2^0/\sigma_2^0)$. Next, we proceed to choosing a sequence of $G \in \mathcal{E}_{k_0}$ as follows: $p_i = p_i^0, \theta_i = \theta_i^0, v_i = v_i^0$ for all $1 \leq i \leq k_0$, and $m_1 = m_1^0 + 1/n, m_2 = m_2^0 - \sigma_2^0/n\sigma_1^0, m_i = m_i^0$ for all $3 \leq i \leq k_0$. The choice of m_1, m_2 is taken to guarantee that $\Delta m_1/\sigma_1^0 + \Delta m_2/\sigma_2^0 = 0$ as we have discussed in Section 7.1.3. Then, we can check that $\sum_{j=1}^2 p_j(m_j^0)^u (\Delta m_j)^v / (\sigma_j^0)^{u+v+1} = 0$ for all odd numbers $u \leq v$ when v is even number or for all even numbers $0 \leq u \leq v$ when v is odd number. From here, the completion of the proof follows in the same way as that of the special case previously described.

8.5 Proofs for auxiliary results

Lemma 8.1. *Let $\{f(x|\theta, \sigma, m), \theta \in \Theta_1, \sigma \in \Theta_2, m \in \Theta_3\}$ be a class of skew normal distribution. Denote $v := \sigma^2$, then*

$$\begin{cases} \frac{\partial^2 f}{\partial \theta^2}(x|\theta, \sigma, m) - 2\frac{\partial f}{\partial v}(x|\theta, \sigma, m) + \frac{m^3 + m}{v} \frac{\partial f}{\partial m}(x|\theta, \sigma, m) = 0. \\ 2m\frac{\partial f}{\partial m}(x|\theta, \sigma, m) + (m^2 + 1)\frac{\partial^2 f}{\partial m^2}(x|\theta, \sigma, m) + 2vm\frac{\partial^2 f}{\partial v \partial m}(x|\theta, \sigma, m) = 0. \end{cases}$$

Proof. Direct calculation yields

$$\begin{aligned} \frac{\partial^2 f}{\partial \theta^2}(x|\theta, \sigma, m) &= \left\{ \left(-\frac{2}{\sqrt{2\pi}\sigma^3} + \frac{2(x-\theta)^2}{\sqrt{2\pi}\sigma^5} \right) \Phi\left(\frac{m(x-\theta)}{\sigma}\right) - \frac{2m(m^2+2)(x-\theta)}{\sqrt{2\pi}\sigma^4} f\left(\frac{m(x-\theta)}{\sigma}\right) \right\} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \\ \frac{\partial f}{\partial v}(x|\theta, \sigma, m) &= \left\{ \left(-\frac{1}{\sqrt{2\pi}\sigma^3} + \frac{(x-\theta)^2}{\sqrt{2\pi}\sigma^5} \right) \Phi\left(\frac{m(x-\theta)}{\sigma}\right) - \frac{m(x-\theta)}{\sqrt{2\pi}\sigma^4} f\left(\frac{m(x-\theta)}{\sigma}\right) \right\} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \\ \frac{\partial f}{\partial m}(x|\theta, \sigma, m) &= \frac{2(x-\theta)}{\sqrt{2\pi}\sigma^2} f\left(\frac{m(x-\theta)}{\sigma}\right) \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \\ \frac{\partial^2 f}{\partial m^2}(x|\theta, \sigma, m) &= \frac{-2m(x-\theta)^3}{\sqrt{2\pi}\sigma^4} f\left(\frac{m(x-\theta)}{\sigma}\right) \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \\ \frac{\partial^2 f}{\partial v \partial m}(x|\theta, \sigma, m) &= \left(-\frac{2(x-\theta)}{\sqrt{2\pi}\sigma^4} + \frac{(m^2+1)(x-\theta)^3}{\sqrt{2\pi}\sigma^6} \right) \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right). \end{aligned}$$

From these equations, we can easily verify the conclusion of our lemma. \square